

SAMPLING AND SELECTION BIAS IN RESEARCH USING DOCUMENTED SKELETAL
COLLECTIONS

By

Rhian Reeves Dunn

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Anthropology – Doctor of Philosophy

2025

ABSTRACT

Bias research in forensic anthropology has focused primarily on the effects of cognitive bias on laboratory analyses. However, bias not only arises from contextual information at the scene, but can also occur during research endeavors, from design and data collection to analysis and interpretation. To date few studies have focused on potential sources of bias in documented skeletal collections, despite their pivotal role in the production of analytical methods in the United States. Therefore, the purpose of this research is to identify potential sources of sample and selection bias in forensic anthropological research using documented skeletal collections.

The three goals of this project address different aspects of sample and selection bias that may be potentially encountered during research development, each addressed in a separate manuscript. The first manuscript uses both demographic and craniometric data from eight United States-based documented skeletal collections to investigate whether specific procurement strategies result in collection specific sample bias. Significant differences between collections were identified in both demographic and craniometric data, indicating the presence of collection-specific sample bias. However, collection-specific sample bias does not obscure patterns of normal human skeletal variation, including sex and population. Therefore, documented skeletal collections remain valid sources of sample data.

The second manuscript investigates whether collection samples are representative of the population of interest in forensic anthropology by comparing craniometric data from eight documented skeletal collection to the craniometric data found in Fordisc 3.1.322. Collections were most representative of case populations well-reflected in their samples, such as European Americans who comprise the majority in every documented skeletal collection. Furthermore, inherent sample bias was identified in Fordisc, as nearly one quarter of individuals originate

from the same data source; inclusion of those data largely impact model performance. Forensic anthropologists are encouraged to continue case submission to the Forensic Anthropology Data Bank to increase the diversity in Fordisc sample sources, bolster reference sample sizes, and increase the number and variety of the demographic populations represented.

The third manuscript uses geostatistical and spatial analyses to investigate whether the physical layout, storage of individuals, and curator involvement at documented human skeletal collections introduce selection bias. Sampling frequency was analyzed using spatial variables, demographic information of individuals within each collection, and interview data, which addressed curator involvement. Selection bias was identified and is mainly attributable to the physical layouts of each collection, including cabinet location and shelf height. However, researcher practices—such as evaluating individuals in chronological order—also contribute to uneven sampling distributions. Researchers need to employ sampling strategies to specifically mitigate the effects of collection layout.

Valid methods are required for expert court testimony and all attempts to mitigate bias are essential to upholding ethical standards in the criminal justice system. This dissertation demonstrates sample and selection biases exist within documented skeletal collections, but these collections remain invaluable resources for research in forensic anthropology. Additionally, this dissertation provides a foundation for future investigations into biases for the many documented skeletal collections not included in this dissertation

To all the individuals who have donated their skeletons, without whom this research would not have been possible.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the enduring support of so many colleagues, friends, and family, to whom I am incredibly grateful.

First and foremost, I would like to express my deepest appreciation to my doctoral committee, Drs. Joseph Hefner, Carolyn Isaac, Gabriel Wrobel, and Maureen Schaefer, for their mentorship, feedback, and continuous encouragement. I have learned so much from each of you, and my dissertation is better for the insight and guidance you have provided. I am especially grateful to my chair and advisor, Dr. Hefner, who has helped me develop as a scholar and researcher and has been nothing short of an exemplary advisor. He has always put his students first, challenged us to think critically about forensic anthropology and life, cultivated a collaborative and welcoming laboratory, and emphasized the importance of work-life balance. I can say with certainty that I would not have completed this dissertation without his guidance, and I hope to emulate his mentorship in my own career.

I would also like to thank the countless curators and collection managers who facilitated my research visits, answered my many questions, and helped develop my dissertation research through thoughtful conversations. Additionally, I sincerely appreciate the financial support from the Corey Endowment and Research Scholars at the College of Social Sciences at Michigan State University, as well as the National Institute of Justice through the 2023 Graduate Research Fellowship. This funding afforded me the great privilege of visiting so many incredible collections and provided me with the time to analyze and write without worry.

To my incredible support system at Michigan State University—thank you. To everyone in the Michigan State University Forensic Anthropology Laboratory, thank you for creating such a welcoming and enjoyable environment. Drs. Isaac and Hefner, I deeply appreciate your

mentorship and the way you have graciously taught me to be a more capable forensic anthropologist. To my fellow graduate students, past and present—your invaluable discussions, encouragement, and much-needed coffee breaks broadened my horizons, kept me motivated, and made each day on campus something to look forward to. To my cohort, thank you for the shared laughter, support, and camaraderie over the years; it has been a pleasure to go on this journey with all of you.

A special thank you is owed to Dr. Dennis Dirkmaat, my master's thesis advisor, for introducing me to forensic anthropology and teaching me to "question everything." I must also thank those at Mercyhurst University and my master's cohort for creating a support system I will lean on for years to come. To Andie and Paige—I would not have made it through my dissertation without both of you by my side. Your friendship and collaboration have been an unshakeable source of strength, and I am endlessly grateful for both.

I owe immense gratitude to my friends and family outside of academia who believed in me and motivated me through the toughest times. I am beyond lucky to have had such a strong support system over the years. To my parents, who instilled in me a strong work ethic, passion, love, and a good sense of humor—though you may not have anticipated I'd be in school for quite this long—I hope this dissertation is a testament to the values you've passed down to me. To my other half, Dr. Megan Dunn, thank you for paving the way and always being there for me, no matter the time or issue—your support has meant the world to me.

And finally, to my husband, Thomas—thank you for everything. I would be lost without your unwavering support, patience, and love. I am blessed to have you in my life as my partner and my greatest source of strength.

TABLE OF CONTENTS

INTRODUCTION	1
DOCUMENTED SKELETAL COLLECTIONS	2
UNDERSTANDING BIAS	5
SUMMARY	8
ORGANIZATION OF DISSERTATION	9
BIBLIOGRAPHY	10
 MANUSCRIPT 1. COLLECTION-SPECIFIC SAMPLE BIAS IN UNITED STATES-BASED DOCUMENTED SKELETAL COLLECTIONS	 13
INTRODUCTION	13
MATERIALS AND METHODS	15
RESULTS	25
DISCUSSION	42
CONCLUSION	48
BIBLIOGRAPHY	49
 MANUSCRIPT 2. REPRESENTATION OF DOCUMENTED SKELETAL COLLECTIONS TO FORENSIC CASEWORK	 53
INTRODUCTION	53
MATERIALS AND METHODS	55
RESULTS	61
DISCUSSION	82
CONCLUSION	86
BIBLIOGRAPHY	87
 MANUSCRIPT 3. GEOSTATISTICAL AND SPATIAL ANALYSIS OF SELECTION BIAS IN DOCUMENTED SKELETAL COLLECTIONS	 90
INTRODUCTION	90
MATERIALS AND METHODS	92
RESULTS	100
DISCUSSION	119
CONCLUSION	122
BIBLIOGRAPHY	124
APPENDIX A: ANONYMOUS SURVEY QUESTIONS	126
APPENDIX B: INTERVIEW QUESTIONS	131
 CONCLUSION	 132

INTRODUCTION

The purpose of this dissertation is to identify potential sources of sample and selection bias in forensic anthropological research using documented skeletal collections. This research focuses on three specific aspects of sample and selection bias using samples from U.S.-based documented skeletal collections. By quantifying bias encountered during the use of these collections, this study addresses a critical gap in current knowledge. Ultimately, this research will improve the reliability and applicability of the methodologies generated using documented skeletal collections.

Bias can arise at any stage of research—from the formulation of research questions and data collection to the analysis and interpretation of that data (Smith and Noble 2014). Failure to address biasing factors can, and usually does, impact a study's reliability and validity (Smith and Noble 2014). In forensic anthropology, the methodologies used to analyze and interpret skeletal data can be presented in criminal court and require validity, repeatability, peer-review, and demonstrable error rates to meet the standards identified in *Daubert v. Dow Pharmaceuticals* (Christensen 2005; Lesciotto and Christensen 2024). To meet these standards, research and methods with bias must be addressed at each stage, including during research design and data collection.

Research on bias in forensic science primarily focuses on cognitive bias, including the effects of contextual information (i.e., external influences which interfere with one's ability to examine, evaluate and judge case data [Dror et al. 2006]), procedures used to make comparisons, previous knowledge concerning some aspect of a case, and emotion levels of the analyst during the decision making process (Cooper and Meterko 2019). Cognitive bias is particularly relevant in forensic anthropology, a field where practitioners are engaged in multiple stages of an investigation, including during the outdoor recovery, the autopsy, laboratory-based skeletal

analyses, and expert witness testimony (Warren et al. 2017). Consequently, research has prioritized understanding how contextual bias introduced during the recovery phase impacts later laboratory analysis and courtroom testimony (c.f., Hartley et al. 2021; Nakhaeizadeh et al. 2014a, 2014b, 2018). While efforts to control contextual biases maintain methodological validity, they largely ignore other forms of bias, like sample and selection bias, which may significantly impact the design of new analytical methods.

Documented skeletal collections—widely used as a primary source of sample data in forensic anthropology—are most likely affected by sample and selection bias. These collections contribute to countless research endeavors, providing samples for a large number of studies developing analytical methods (Campanacho et al. 2021); therefore, there is an imperative need to identify and address potential sources of bias encountered during research using those collections. All forms of bias must be systematically examined to develop effective countermeasures.

DOCUMENTED SKELETAL COLLECTIONS

Documented skeletal collections are aggregations of skeletal remains from individuals with associated demographic information (e.g., age, sex, social race, stature [Campanacho et al. 2021]). Within the U.S., many of the early documented skeletal collections were associated with medical schools, where early physical anthropologists working as anatomists would retain skeletal remains from dissection cadavers. These collections provided an avenue to study normal and pathological human skeletal variation (Muller et al. 2017). Notable historic collections include the George S. Huntington Anatomical Skeletal Collection, the Hamann-Todd Human Osteological Collection, the Robert J. Terry Anatomical Collection, and the William Montague Cobb Collection of Human Skeletal Remains (Muller et al. 2017).

These collections played a key role in the development of biological and forensic

anthropology, as they were used in nearly every project developing early analytical methods (Campanacho et al. 2021; Muller et al. 2017). However, these collections have been criticized in recent years due to less than ideal, and at times unethical, acquisition methods for the cadavers (Sholts 2024). Prior to the Uniform Anatomical Gift Act in 1968 (Sadler et al. 1968), cadavers were primarily sourced from unclaimed individuals; in some states families were given as little as twenty-four hours to claim deceased individuals (Muller et al. 2017). This framework ensured medical schools had adequate material for instruction, but it disproportionately targeted and impacted impoverished groups, including recent immigrants and the minoritized. The demographic composition of the Terry Collection demonstrates this disparity, which is over 50% African American despite African Americans representing less than 10% of the living St. Louis population when the Terry Collection was acquiring skeletons (de la Cova 2019). Likewise, immigrants, including many Germans, Irish, and Italian, make up over 50% of the Huntington Collection (Muller et al. 2017). In Cleveland, regional mortuaries, city hospitals, charity institutions, and the Cleveland Workhouse were required by law to notify T. Wingate Todd, the second director of the Hamann-Todd Collection, when they had unclaimed decedents, ensuring a steady supply of cadavers (Muller et al. 2017). Ultimately, poor houses, mental institutions, hospitals, and long-term care facilities provided the lion's share of unclaimed bodies used in dissection rooms (de la Cova 2021; Muller et al. 2017).

Many anthropologists also argue historic collections no longer represent the modern U.S. population since they primarily house individuals from low socioeconomic status groups with birth years prior to 1900 (Jantz and Jantz 2000; Jantz and Jantz 2016; Jantz and Moore-Jansen 1987; Jantz et al. 2016). In response to shifting genetic and environmental factors over the last century, including changes in average levels of nutrition, living conditions, medical advancements, and

socioeconomic status, secular change has been documented in various skeletal proportions (Kilroy et al. 2020). For example, the U.S. population has increased in both limb length and overall stature (Jantz and Jantz 1999). As such, stature estimation methods developed using historic collections consistently underestimate height and are no longer appropriate for modern populations (Jantz and Jantz 1999; Ousley 2012). Furthermore, secular change has been documented in cranial dimensions; faces have become more narrow, and have higher, longer vaults (Jantz and Jantz 2000). Similarly, nonmetric traits used in population affinity estimation and sex estimation also have demonstrable levels of secular change (Kilroy et al. 2020; Klales 2016). Although not all of the methods used by forensic anthropologists are affected by secular changes (Klales 2016), others, such as stature estimation, can be affected, so a level of caution is warranted when using methods on modern populations that were developed using historic collections.

To address limitations in historical collections over the past several decades, modern documented skeletal collections have been started (Campanacho et al. 2022). These collections 1) contain individuals with more recent birth years to ensure they are more representative of the current demography; 2) prioritize the collection of detailed demographic information; and, 3) maintain strict donation guidelines to ensure ethical acquisition (Campanacho et al. 2022). These practices are invaluable for method development and research in forensic anthropology using these collections.

However, these modern collections also face challenges, including limited donor diversity and generally poor geographic representation (Winburn et al. 2022). Most documented collections report an overrepresentation of European Americans (e.g., George et al. 2022; Gocha et al. 2022; Winburn et al. 2022), which Winburn and colleagues (2022) attribute to body donation avoidance by minorities and immigrants in response to historic victimization in the name of science.

Collections also report lower percentages of females and younger individuals (e.g., George et al. 2022; Gocha et al. 2022; Winburn 2022). These shortcomings ultimately lead to gaps in demographic representativeness (Campanacho et al. 2021). While efforts are made to prioritize donations to address imbalances, limited population demographics severely impact research endeavors, including validation efforts, with the potential to hinder future identifications in forensic anthropological casework (Campanacho et al. 2021).

The lack of adequate geographic representation also poses a challenge for modern collections. Collections generally reflect the surrounding demography rather than the broader U.S. population from influencing factors like state laws governing body donation, demographics of individuals willing to donate, and financial constraints surrounding the transportation of human remains (Campanacho et al. 2021). This disproportionate representation within a region may impact the applicability of research findings to the greater U.S. population. Understanding these limitations as potential sources of bias is crucial for forensic anthropological research since these biases can impact the development and validation of methods and the interpretation of skeletal data in research and casework.

UNDERSTANDING BIAS

Bias is an inherent challenge during any research activity, generally arising when systematic errors affect the accuracy and validity of findings (Smith and Noble 2014). Forensic anthropologists rely heavily on the visual assessment of skeletal remains for their analysis, so the primary focus of previous bias research is the impact of cognitive bias (Nakhaeizadeh et al. 2014a, 2014b, 2020) and objectivity (Ubelaker 2021; Warren 2019; Winburn 2018). However, this dissertation focuses on biases encountered during the early stages of research, while the researcher is developing a study design and collecting data. Below, sample and selection bias are explored

and their relevance contextualized. These concepts provide a lens through which this dissertation quantifies the presence and impact of bias during research design and development using documented skeletal collections, particularly concerning their impact on forensic anthropological research.

Sample Bias

Sample bias occurs when a study sample is systematically different from the population of interest (Sica 2006). Sample bias can result from sample selection or a limited availability of potential participants for a study (Smith and Noble 2014). Because an inadequate sample impacts the significance of study findings, researchers are encouraged to employ random sampling strategies and conduct power analyses to ensure a sufficient and representative sample (Shorten and Moorley 2014). However, in forensic anthropology, sample bias is a potential issue because of the dependence on documented skeletal collections. These collections are not randomly sampled from the living population—they are constrained by the willingness of individuals to donate their bodies (Albanese 2018). The result in many collections across the U.S. is an overrepresentation of older, European American males, the demographic most likely to engage in body donation (Campanacho et al. 2021). Compared to state and national census demographic distributions, documented skeletal collections are systematically different (Winburn et al. 2022). However, forensic anthropologists need skeletal remains for their research. As such, sample bias needs to be better understood to develop mitigation strategies.

Although all collections exhibit some level of sample bias, the extent varies between collections. Collections generally receive donations from individuals in local communities (Campanacho et al. 2021). Each collection has different requirements and strategies for accepting donations (Albanese 2018). For example, the University of Tennessee, Knoxville, Donated Skeletal Collection is actively working to balance unequal sex ratios by prioritizing female donors

(Campanacho et al. 2021). In contrast, the Mann-Labrash Osteological Collection in Honolulu, Hawaii prioritizes donations from diverse population groups representative of Hawaii (Mann et al. 2020). These collection-specific factors and decisions shape donor populations and have the potential to create collection-specific sample bias.

Representativeness refers to the degree to which a sample reflects the spectrum of characteristics in a target population (Sica 2006). Although random sampling ensures all potential participants have an equal chance of inclusion in a study, certain research questions require adequate representation from multiple groups (for example, males and females or African Americans and European Americans). In those cases, sampling strategies like stratified random sampling, where individuals are chosen randomly from a set of predetermined groups, ensures more balanced inclusion (Sica 2006). As discussed previously, certain groups, such as European American males, are overrepresented in documented collections; while others, such as females and minorities, are underrepresented in those collections (Campanacho et al. 2021). However, the primary population of interest for forensic anthropologists is not the living demographic but rather the population that makes up their case load (Franklin and Marks 2021).

If documented collections are unrepresentative of forensic anthropological casework, strategies for expanding the demographic diversity of the collections used to develop methods, procedures, and protocols need to be considered. This is a critical step to ensure the validity and applicability of forensic anthropological methods and to enable more equitable and accurate identifications in forensic anthropology.

Selection Bias

Selection bias occurs when the selection process or inclusion criteria for a study results in a sample that is not representative of the population of interest (Smith and Noble 2014). To mitigate selection bias, researchers must identify target groups in advance and utilize objective

sampling strategies (Smith 2019). While the existing literature on selection bias generally focuses on issues encountered in qualitative studies, like participation bias or self-selection bias (Sica 2006), sampling from documented skeletal collections differs significantly from these other studies since it does not involve interactions with living participants. However, researchers working with documented skeletal collections may still encounter selection biasing factors unique to those collections.

Documented skeletal collections often have specific research protocols and logistical constraints that significantly influence sampling strategies (e.g., Body Donation 2024). These constraints may involve physical considerations, such as the need for ladders to access individuals stored on tall shelves, or logistical challenges, such as collections distributed across multiple locations. Curators or collection managers at each facility may offer guidance or advice, inadvertently shaping sample selection. While these factors are frequently overlooked, they have the potential to introduce bias by affecting which individuals within a collection are sampled.

The extent to which selection bias exists in research using documented skeletal collections has not been fully explored. If individuals housed in a more easily accessible area are more likely to be sampled and those in more inconvenient locations are overlooked, sampling practices may inadvertently skew research findings and impact measures of human variation. To address this issue, forensic anthropologists must evaluate current sampling practices and, if necessary, develop new strategies tailored to the challenges of working with skeletal collections.

SUMMARY

Overall, this dissertation will contribute to the anthropological understanding of bias in research using documented skeletal collections and how that bias may impact their validity. Through quantitative analysis of skeletal and geospatial data, I examine how sample and selection

biases influence our understanding of human skeletal variation and qualify the continued use of collections in forensic anthropological research. These analyses address challenges posed by bias and identify areas where forensic anthropologists must increase mitigation efforts. By addressing this gap in current knowledge, this dissertation informs the reliability and validity of current forensic anthropological methods developed using documented skeletal collections.

ORGANIZATION OF DISSERTATION

Three separate aspects of research bias will be examined, including: 1) collection-specific sample bias in documented skeletal collections, 2) the representativeness of documented skeletal collections to forensic anthropological casework, and 3) selection bias in research sampling from documented skeletal collections. A final chapter synthesizes these, interpreting the impact of research bias and discussing greater implications to the field of forensic anthropology.

Manuscript One (*“Collection-Specific Sample Bias in United States-based Documented Skeletal Collections”*) investigates the presence of collection-specific sample bias using data from eight modern U.S.-based documented skeletal collections. Manuscript Two (*“Representation of Documented Skeletal Collections to Forensic Casework”*) focuses on whether documented skeletal collections are representative of forensic anthropology casework using data from documented skeletal collections and the Forensic Anthropology Data Bank. Finally, Manuscript Three (*“Geostatistical and Spatial Analysis of Selection Bias in Documented Skeletal Collections”*) examines the presence of selection bias in studies using documented skeletal collection through geostatistical and spatial analyses of sampling frequencies.

BIBLIOGRAPHY

- Albanese, J. 2018. Strategies for dealing with bias in identified reference collections and implications for research in the 21st century. In: C. Y. Henderson & F. A. Cardoso, eds. *Identified skeletal collections: The testing ground of anthropology?* Oxford, UK: Archaeopress Publishing LTD. pp. 59-82.
- Body Donation. 2024. Forensic Anthropology Center, University of Tennessee, Knoxville. Retrieved October 13, 2024, from <https://fac.utk.edu/body-donation/>
- Campanacho, V., Ales Cardoso, F., & Ubelaker, D. H. 2021. Documented skeletal collections and their importance in forensic anthropology in the United States. *Forensic Sciences* 1:228-239. <https://doi.org/10.3390/forensicsci1030021>
- Christensen, Angi M. 2005. Testing the Reliability of Frontal Sinuses in Positive Identification. *Journal of Forensic Sciences* 50(1): 1–5. <https://doi.org/10.1520/jfs2004145>
- Christensen, A. M., & Passalacqua, N. V. 2018. *A laboratory manual for forensic anthropology*. London, UK: Academic Press. <https://doi.org/10.1016/C2016-0-03295-3>
- Cooper, G. S., & Meterko, V. 2019. Cognitive Bias Research in Forensic Science: A Systematic Review. *Forensic Science International* 297: 35-46. <https://doi.org/10.1016/j.forsciint.2019.01.016>
- de la Cova, C. 2019. Marginalized bodies and the construction of the Robert J. Terry Collection: A promised land lost. In: M. L. Mant & A. J. Holland, eds. *Bioarchaeology of Marginalized People*. London, UK: Springer Inc. pp. 133-155. <https://doi.org/10.1016/C2017-0-02300-5>
- de la Cova, C. 2021. Making silenced voices speak: Restoring neglected and ignored identities in anatomical collections. In: C. M. Cheverko, J. R. Prince-Buitenhuis & M. Hubbe, eds. *Theoretical perspectives in bioarchaeology*. New York, NY: Routledge Taylor & Francis Group. pp. 150-169.
- Dror, I. E., Charlton, D., & Péron, A. E. 2006. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International* 156:74-78.
- Franklin, D., & Marks, M. K. 2021. The professional practice of forensic anthropology: Contemporary developments and cross-disciplinary applications. *WIREs Forensic Science* 4(2): e1442.
- George, R. L., Zejdlik, K., Messer, D. L., & N. V. Passalacqua. 2022. The John A. Williams Human Skeletal Collection at Wester Carolina University. *Forensic Sciences* 2:362-370. <https://doi.org/10.3390/forensicsci2020026>
- Gocha, T. P., Mavroudas, S. R., & Wescott, D. J. 2022. The Texas State Donated Skeletal Collection at the Forensic Anthropology Center at Texas State. *Forensic Sciences* 2(1):7-19. <https://doi.org/10.3390/forensicsci2010002>

- Hartley, S., Windburn, A. P., & Dror, I. E. 2021. Metric Forensic Anthropology Decisions: Reliability and Biasability of Sectioning-Point-Based Sex Estimates. *Journal of Forensic Sciences* 67(1): 68-79. <https://doi.org/10.1111/1556-4029.14931>
- Jantz, L. M., & Jantz, R. L. 1999. Secular change in long bone length and proportion in the United States, 1800-1970. *American Journal of Physical Anthropology*, 110, 57-67.
- Jantz, L. M., & Jantz, R. L. 2000. Secular change in cranial facial morphology. *American Journal of Human Biology*, 12(3), 327-338. [https://doi.org/10.1002/\(SICI\)1520-6300\(200005/06\)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1520-6300(200005/06)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1)
- Jantz, R. L., & Meadows Jantz, L. 2016. The remarkable change in Euro-American cranial shape and size. *Human Biology* 88(1): 56-64.
- Jantz, R. L., Meadows Jantz, L., & Devlin, J. L. 2016. Secular changes in the postcranial skeleton of American Whites. *Human Biology*, 88(1), 65-75. <https://doi.org/10.13110/humanbiology.88.1.0065>
- Jantz R. L., & Moore-Jansen, P. H. 1987. *Final report to the National Institute of Justice: Grant No. 85-IJ-CX-0021. Department of Anthropology*, University of Tennessee, Knoxville.
- Kilroy, G. S., Tallman, S. D., & DiGangi, E. A. 2020. Secular change in morphological cranial and mandibular trait frequencies in European Americans born 1824-1987. *American Journal of Biological Anthropology* 173(3):589-605.
- Klales, A. R. 2016. Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Science* 61(2): 295-301.
- Lesciotto, K. M., & Christensen, A. M. 2024. The over-citation of Daubert in forensic anthropology. *Journal of Forensic Sciences* 69(1): 9-17.
- Mann, R. W., Labrash, S., & Lozanoff, S. 2020. A new osteological resource at the John A. Burn's School of Medicine. *Hawai'i Journal of Health & Social Welfare* 79(6): 202-203.
- Muller, J. L., Pearlstein, K. E., & de la Cova, C. 2017. Dissection and documented skeletal collections: Embodiments of legalized inequality. In: K. C. Nystrom, ed. *The bioarchaeology of dissection and autopsy in the United States*. New Paltz, NY: Springer. pp. 185-201.
- Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. 2014a. Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias. *Science & Justice* 54(3): 208-214. <https://doi.org/10.1016/j.scijus.2013.11.003>
- Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. 2015. The emergence of cognitive bias in forensic science and criminal investigations. *British Journal of American Legal Studies* 4: 527-554.

- Nakhaeizadeh, S, Hanson, I., & Dozzi, N. 2014b. The Power of Contextual Effects in Forensic Anthropology: A Study of Biasability in the Visual Interpretations of Trauma Analysis on Skeletal Remains. *Journal of Forensic Sciences* 59(5): 1177-83. <https://doi.org/10.1111/1556-4029.12473>.
- Nakhaeizadeh, S, Morgan, R. M., Rando, C., & Dror, I. E. 2018. Cascading Bias of Initial Exposure to Information at the Crime Scene to Subsequent Evaluation of Skeletal Remains. *Journal of Forensic Sciences* 63(2): 403-11. <https://doi.org/10.1111/1556-4029.13569>
- Ousley, S. D. 2012. Estimating stature. In: D. C. Dirkmaat, ed. *Companion to forensic anthropology*. West Sussex, UK: Blackwell Publishing Ltd. pp. 330-334.
- Ross, A. H., Ubelaker, D. H., & Kimmerle, E. H. 2011. Implications of dimorphism, population variation, and secular change in estimating population affinity in the Iberian Peninsula. *Forensic Science International* 206(1-3): 214.e1-5.
- Sadler, A. M., Sadler, B. L., Stason, E. B. 1968. The Uniform Anatomical Gift Act: A model for reform. *JAMA* 206(1):2501-2506. doi: 10.1001/jama.1968.03150110049007
- Sholts, S. B. 2024. "To honor and remember": An ethical awakening to African American remains in museums. *American Journal of Biological Anthropology*, Early View:e24943. <https://doi.10.1002/ajpa.24943>
- Shorten, A., & Moorley, C. 2014. Selecting the sample. *Evidence Based Nursing* 17(2): 32-33.
- Sica, G. T. 2006. Bias in research studies. *Radiology* 237(3): 780-789.
- Smith, R. 2019. Living with observational data in biological anthropology. *American Journal of Biological Anthropology* 169:591-598.
- Smith, J., & Noble, H. 2014. Bias in Research. *Evidence Based Nursing* 17(4): 100-101. <https://doi.org/10.1136/eb-2014-101946>
- Ubelaker DH. 2021. Research integrity in forensic anthropology. *Forensic Sciences Research* 6(4): 285-291. <https://doi.org/10.1080/20961790.2021.1963515>
- Warren, M. W., Friend, A., & Stock, M. K. 2017. Navigating Cognitive Bias in Forensic Anthropology. In: . Clifford Boyd and Donna C. Boyd, eds. *Forensic Anthropology: Theoretical Framework and Scientific Basis*. New Jersey: John Wiley & Sons Ltd. pp. 39-51. <http://dx.doi.org/10.1002/9781119226529.ch3>.
- Winburn, A. P. (2018). Subjectivity with a capital S? Issues of objectivity in forensic anthropology. In: C. C. Boyd & D. C. Boyd, editors. *Forensic anthropology: Theoretical framework and scientific bias*. Hoboken, NJ: John Wiley & Sons LTD. pp. 19-38.
- Winburn, A. P., Jennings, A. L., Steadman, D. W., & DiGangi, E. A. 2022. Ancestral diversity in skeletal collections: Perspectives on African American body donation. *Forensic Anthropology* 5(2):141-152. <https://doi.org/10.5744/fa.2020.1023>

MANUSCRIPT 1. COLLECTION-SPECIFIC SAMPLE BIAS IN UNITED STATES- BASED DOCUMENTED SKELETAL COLLECTIONS

INTRODUCTION

Documented skeletal collections are an integral component of research in biological anthropology, providing a testing ground for the creation and validation of methods utilized for skeletal analysis—the very bread and butter of our discipline (Campanacho et al. 2021, de la Cova 2021; Henderson 2018). Considering their importance to research, we need to understand if these collections are representative of the population or if they are intrinsically biased.

Proper research protocols emphasize the importance of a representative sample to ensure reliability and validity in study results (Sharma 2017; Singh and Masuka 2014). Ideally, all individuals in a population would be included in a study; however, if this is impractical, systematic sampling is recommended to ensure unbiased results and adequate representation (Smith and Noble 2014; Smith 2019). In biological anthropology, systematic sampling may not be possible since we cannot randomly sample skeletons from a living population. Instead, we must use available sources, including documented skeletal collections, to conduct our research.

Documented skeletal collections are not representative of the living population because individuals do not have an equal probability of dying, except in unique cases like massacres or catastrophes (Albanese 2018). Similarly, not all decedents have an equal likelihood of being included in a documented skeletal collection. Before the legalization of body donation and the enactment of the Uniform Anatomical Gift Act (UAGA) in 1968, most individuals in documented collections were those unclaimed by family members. This resulted in an overrepresentation of individuals from lower socioeconomic status (SES) groups of predominately immigrant and minoritized populations (de la Cova 2021; Sadler et al. 1968). For instance, African Americans

make up over half of the entire Robert J. Terry Collection yet represented less than 10% of St. Louis's living population in the 1930s and 40s (de la Cova 2019). Today, willed body donation is legal and less socially stigmatized; however, there are individuals who strongly oppose body donation for numerous reasons, including skepticism concerning body disfiguration, religious restriction, and the historic victimization of donors by researchers (Winburn et al. 2020). The result has been that older European American males represent the highest donation rate, a fact clearly reflected in demographic compositions of United States-based documented collections (George et al. 2022; Gocha et al. 2022; Komar and Grivas 2008; Winburn et al. 2020). These factors have culminated in biased historic and modern collections, which have demographic compositions that do not reflect either state or national census populations (Winburn et al. 2020).

Even among potential donors, there is an unequal likelihood of donation to any specific collection. Documented skeletal collections are often affiliated with medical schools, medical examiner's offices, and/or outdoor decomposition facilities whose locations and professional networks impact which individuals are available for donation. Each collection also sets specific criteria for accepting potential donations. Factors like space limitations, resource constraints for processing donations, and restrictions based on donor characteristics (e.g., height, weight, active communicable diseases, unclaimed decedent, estranged next-of-kin) influence a curator's decisions (Armelli et al. 2022). The individual procurement strategies of curators, such as targeting or avoiding specific demographic variables, can also influence the composition of a collection. An historic example of such targeting is Trotter's attempt to balance the uneven sex distribution by prioritizing female donors in the Terry Collection (de la Cova 2021; Sharman and Albanese 2018). Ultimately, these factors impact which donors are included in a collection, resulting in unique samples—or, collection-specific sample bias.

The question is no longer whether collection-specific sample bias exists—we know it does. The real question is whether this bias produces artificial "populations" within collections that mask genuine patterns of human skeletal variation. For instance, secular change studies identified significant changes in skeletal proportions between historic and modern collections due to shifting genetic and environmental components over time (Jantz and Jantz, 2000, 2016; Langley et al. 2016; Moore-Jansen 1989; Spradley 2008; Spradley et al. 2016). Many researchers argue these findings prove historic collections are biased and should not be used for modern research (Sherman and Albanese 2018). On the other hand, Albanese (2003, 2018) argues the impact of sample bias is not limited to historic collections and can be documented across modern collections as well. If sample bias can impact our understanding of aggregate skeletal morphology, are documented skeletal collections valid and reliable sources for understanding population diversity, or do observed differences simply reflect collection sample structures?

To assess collection-specific sample bias and how it can impact our understanding of the variation in skeletal morphology, documented skeletal collections from similar temporal periods and geographic regions with comparable population histories should be compared. Examining documented collections from similar contexts can reveal how collection-specific sample bias impacts skeletal morphology and if that bias obscures true patterns of human skeletal variation, such as population differences. This study investigates the presence of collection-specific sample bias and the extent that bias may impact, or overwhelm, normal human skeletal variation, using data from eight modern documented skeletal collections from the United States.

MATERIALS AND METHODS

Samples

Collection-specific sample bias was assessed using craniometric data for 1161 individuals from eight documented skeletal collections located across the United States (Table 1.1).

Craniometric data are valid proxies of genetic markers due to measurable differences in geographic patterning, sexual dimorphism, plasticity, and secular change (e.g., Jantz and Jantz, 2000; Relethford, 2009; Sparks and Jantz, 2002), as well as anthropometric patterns not attributable to general human variation (Gordon and Bradtmiller 1992; Utermohle and Zegura 1982; Utermohle et al. 1983). Previous research investigating sample bias in documented skeletal collections also used metric data (Albanese 2003; Moore-Jansen 1989). Craniometric data for this study comprise 78 standard landmarks collected using a Microscribe® G digitizer and the 3Skull software program (Fleischman and Crowder 2019; Ousley 2014). All data were collected by the first author (RRD) to govern observer error between collection samples. To contextualize the data, demographic variables were requested from the individual institutions. These variables include: age, birthplace, birth year, cause of death, height, location at time of death, occupation, sex, social race, SES, type of donation (e.g., self-donation, next-of-kin donation), and year of death.

Documented skeletal collections were selected for inclusion based on their use in biological anthropological research, the collection's standing as a modern skeletal collection (i.e., established after the UAGA and still accepting donations; Campanacho et al. 2021), the geographic location, and whether access is permitted to visiting researchers (e.g., the Hamann-Todd Human Osteological Collection is currently closed to researchers for museum renovations). Ultimately, eight collections were included. Background information about each collection is provided below and presented in the order in which data were collected for this study.

Maxwell Museum's Documented Skeletal Collection (MMDSC)—Located in the Maxwell Museum of Anthropology at the University of New Mexico (UNM) in Albuquerque, New Mexico, the MMDSC was established in 1978 and contains 314 individuals (as of October 2023). The collection primarily includes residents of New Mexico at the time of their death and all donors

were obtained through self-donation, next-of-kin donation, and/or transfers from the Department of Anatomy at UNM and the Office of Medical Examiner (Komar and Grivas 2008). However, many of the individuals who were transferred have no or limited associated donation information (Komar and Grivas 2008). The MMDSC is primarily males (59%), middle-aged individuals (mean age of 67), and European Americans (91.9%), followed by Hispanic and African American. For this study, craniometric data were collected from all adult crania unaffected by extensive pathology or trauma and having associated demographic information. The final MMDSC sample includes data for 246 individuals (females = 100; males = 146; see Table 1.1).

The University of Tennessee, Knoxville Donated Skeletal Collection (UTK DSC)—Located in the Department of Anthropology at the University of Tennessee, Knoxville (UTK) in Knoxville, Tennessee, the UTK DSC is associated with the Anthropology Research Facility (established in 1981) and contains over 1800 individuals (as of April 2024). The UTK DSC historically accepted donations from medical examiner offices (Bass and Jefferson 2003), but currently only accepts donations from donors or donor families, free of cost and with transportation provided within 100 miles of UTK (Body Donation 2024). The demographic composition of the UTK DSC is primarily males (64%), middle-aged individuals (mean age of approximately 63), and European Americans (93%), followed by African American, Hispanic, Multiple¹, Native American, and Asian American (Age, Sex and Ancestry Distribution 2024; Winburn et al. 2020). For this study, the current curator, Dr. Dawnie Steadman, provided a randomized list of 295 individuals with associated demographic information. Craniometric data were collected from 220 of these individuals (female = 109; male = 111; see Table 1.1). Of note, this sample only includes

¹ A population of “Multiple” was provided if an individual had multiple associated populations/ancestries, such as “European American and Hispanic”. This population category was used by multiple collections in their sample demographic summaries (e.g., Age, Sex and Ancestry Distribution 2024; Gocha et al. 2022).

individuals donated after 2005, because donation forms were updated to include socioeconomic status information starting in 2006.

The Texas State Donated Skeletal Collection (TXSTDSC)—Located in the Forensic Anthropology Center at Texas State (FACTS) in San Marcos, Texas, the TXSTDSC is associated with the Forensic Anthropology Research Facility (FARF) established in 2008 and contains over 700 individuals (Gocha et al. 2022). All donations are from “living” donors (i.e., self-donation) or next-of-kin donations. FACTS does not accept unclaimed decedents or next-of-kin donations from estranged family members (Gocha et al. 2022). Similar to the UTK DSC, FACTS has a limited pick-up radius and can only provide free transportation up to 100 miles from the FARF (Gocha et al. 2022). The demographic composition of the collection is primarily male (58%), middle-aged (mean age of 66), and European American (90%), followed by Hispanic, African American, Multiple, Native American, Asian, and Middle Eastern (Gocha et al. 2022). For this study, the current curator, Dr. Daniel Wescott, provided a random sample of 365 individuals. Craniometric data were collected from 250 individuals (females = 107; males = 143; see Table 1.1).

John A. William’s Documented Human Skeletal Collection (JAWDHSC)—Located in the Department of Anthropology and Sociology at Western Carolina University in Cullowhee, North Carolina, the JAWDHSC was established in 2003 and contains 129 individuals, with a relatively balanced sex distribution, predominately representing European Americans (91% as of May 2024; George et al. 2022). Apart from a few early donations, all donors are self-donations or next-of-kin donations (George et al. 2022). Of note, the JAWDHSC is associated with the Forensic Osteology Research Station and is only a two hour and thirty-minute drive from UTK’s Anthropology Research Facility, which highlights the similar geographic region and research efforts between the two institutions. Demographic information was not available for all individuals, so a sample of

118 individuals (female = 59; male = 59; see Table 1.1) was utilized.

Western Michigan University Homer Stryker M.D. School of Medicine Skeletal Teaching and Research Series (WMed STARS)—Located at Western Michigan University Homer Stryker M.D. School of Medicine (WMed) in Kalamazoo, Michigan, the WMed STARS was established in 2014 and currently houses 53 individuals. A small number of these individuals are currently on loan to the outdoor decomposition facility at Northern Michigan University (as of May 2024). The WMed STARS contains only individuals who have opted for permanent skeletal donation through WMed’s Body Donation Program. The series primarily comprises males and European Americans (72% and 66%, respectfully, as of May 2024). However, demographic information was not available for all individuals, so we retained a sample of 40 individuals (female = 12; male = 28; see Table 1.1).

Michigan State University Forensic Anthropology Laboratory Donated Skeletal Collection (MSUFAL DSC)—Located in the Department of Anthropology at Michigan State University in East Lansing, Michigan, the MSUFAL DSC began accepting donations in 1996 and now contains 42 individuals, including one fetal individual. Historically, the MSUFAL DSC accepted donations from local medical examiners, but has stopped accepting donations that are not self or next-of-kin donations (C. Isaac, personal communication, 2022); however, many donations are still coordinated through contacts at local medical examiner offices. As of September 2024, the demographic composition of the collection is primarily males (67%) and European Americans (75%), followed by African American (23%). For this study, craniometric data were collected from all adult crania not affected by extensive pathology or trauma with known demographic information, resulting in a sample of 37 individuals (females = 12; males = 25; see Table 1.1).

Mann-Labrash Osteological Collection (MLOC)—Located in the Department of Anatomy,

Biochemistry, and Physiology at the John A. Burn's School of Medicine in Honolulu, Hawaii, the MLOC contains over 350 individuals, the first donated in 1974 (Mann et al. 2020). Approximately 50 complete skeletons are available for teaching and research; the remainder only include select elements, such as the skull (as of September 2024). Individuals within MLOC comprise diverse ancestries and ethnicities often not encountered in other United States-based documented skeletal collections, including many individuals of Asian and/or Pacific Island descent (Mann et al. 2020). One goal of the current curators is to target individuals reflecting the diverse living population of Hawai'i and individuals with unique trauma or pathological conditions (R. Mann, personal communication, 2024). All individuals housed within the collection are permanent donations through the Willed Body Program, following informed consent obtained from the donor or the donor's family (Mann et al. 2020). For this study, 100 individuals were randomly selected from the collection (female = 44; male = 56; see Table 1.1).

Southeast Texas Applied Forensic Science Facility Donated Skeletal Collection (STAFS DSC)—Located at the Southeast Texas Applied Forensic Science Facility in Huntsville, Texas, the STAFS DSC was established in 2009 and contains over 450 individuals (STAFS Research 2024). STAFS only accepts living donors (i.e., individuals who have pre-registered for donation upon their death [Gocha et al. 2022]) or next-of-kin donations and is primarily male (67%), over 66 years of age (44%), and European American (84%), followed by African Americans, Other², Hispanic, and Asian (STAFS Research 2024). Of note, STAFS is approximately three hours from the TXSTDSC and both collections are associated with an outdoor decomposition facility. Due to space constraints at STAFS, the collection is divided between STAFS and collection space on the Sam Houston State University main campus. For this study, 150 individuals were randomly

² “Other” is listed as one of the ancestry categories on the STAFS DSC website, but no definition is provided.

selected from both locations (female = 54; male = 96; see Table 1.1).

Table 1.1. Skeletal samples used in this study, broken down by Collection, population, and sex.

Sample (by Collection and Population)*	Male (n)	Female (n)
Maxwell Museum's Documented Skeletal Collection (n = 246)		
African American	5	2
European American	132	94
Hispanic	9	3
Multiple	0	1
University of Tennessee, Knoxville, Donated Skeletal Collection (n = 220)		
African American	0	1
European American	110	105
Hispanic	0	1
Multiple	1	2
Texas State Donated Skeletal Collection (n = 250)		
African American	0	3
European American	129	94
Hispanic	7	5
Native American	0	1
Multiple	7	4
John A. William's Documented Human Skeletal Collection (n = 118)		
African American	1	0
European American	58	59
WMed Skeletal Teaching and Research Series (n = 40)		
African American	0	1
European American	26	9
Native American	1	1
Multiple	1	1
MSUFAL Donated Skeletal Collection (n = 37)		
African American	5	4
European American	20	8
Mann-Labrash Osteological Collection (n = 100)		
African American	3	0
Asian American and Pacific Islander	15	13
European American	32	24
Multiple	6	7
STAFS Donated Skeletal Collection (n = 150)		
African American	7	1
Asian American	1	0
European American	81	51
Hispanic	6	2
Native American	1	0
Total	664	497

* "Multiple" includes those individuals identifying to more than one population affinity.

Statistical Methods

All statistical analyses were conducted in R (R Core Team 2024) and open-source code is available from the authors. To identify sources of collection-specific sample bias and to guide interpretation of any patterns revealed during analysis of craniometric data a series of one-way analyses of variance (ANOVAs) were performed on the demographic variables from each of the eight collections. For each significant result, pairwise comparisons were assessed using a post hoc Tukey HSD test. Demographic variables investigated include sex, population, age, height, birth year, year of death, birth state, state at time of death, and SES. These variables were not available for every collection or every individual, so sample sizes reflect these filtered analyses.

Twenty-nine interlandmark distances were selected for craniometric analysis to match the data commonly used in other forensic anthropological studies and to minimize missing values (Table 1.2). However, some missing data remains. Because complete observations are necessary for multiple analyses, missing values were imputed, by variable, using the ‘*mice*’ package and the predictive mean matching approach, an appropriate method for imputing continuous craniometric data (Azur et al. 2011; Kamnikar et al. 2021; van Buuren and Groothuis-Oudshoorn 2011). Following imputation, boxplots, pairwise plots, and Cook’s distance were used to detect potential outliers (Kamnikar et al. 2021; Techataweewan et al. 2021) and to assess the goodness-of-fit of the imputation process. Following imputation, the craniometric data were standardized by scaling and centering each variable (adjusting the mean to 0 and the standard deviation to 1), ensuring comparability and equal contribution of all variables to the model (Kamnikar et al. 2021).

Table 1.2. Interlandmark distances (and abbreviations) used throughout the analyses.

Abbreviation	Measurement	Abbreviation	Measurement
<i>GOL</i>	Cranial length	<i>OBH</i>	Orbital height
<i>BNL</i>	Cranial base length	<i>OBH</i>	Orbital breadth
<i>BBH</i>	Basion-bregma height	<i>DKB</i>	Interorbital breadth
<i>XCB</i>	Cranial breadth	<i>WNB</i>	Simotic chord
<i>XFB</i>	Maximum frontal breadth	<i>ZMB</i>	Bimaxillary breadth
<i>WFB</i>	Minimum frontal breadth	<i>EKB</i>	Biorbital breadth
<i>ZYB</i>	Bizygomatic breadth	<i>FRC</i>	Frontal chord
<i>AUB</i>	Biauricular breadth	<i>PAC</i>	Parietal chord
<i>ASB</i>	Biasterrionic breadth	<i>OCC</i>	Occipital chord
<i>BPL</i>	Basion-prosthion length	<i>FOL</i>	Foramen magnum length
<i>NPH</i>	Nasion-prosthion length	<i>FOB</i>	Foramen magnum breadth
<i>NLH</i>	Nasal height	<i>MOW</i>	Mid-orbital width
<i>JUB</i>	Bijugal breadth	<i>UFBR</i>	Upper facial breadth
<i>NLB</i>	Nasal breadth	<i>UFHT</i>	Upper facial height
<i>MAL</i>	Max. Alveolar length		

Inter- and intraobserver error tests were performed on the craniometric dataset. For interobserver error testing, interlandmark distances from 41 individuals housed in the MMDSC were provided by the Forensic Anthropology Data Bank (Jantz 2019). For intraobserver error, five individuals were randomly selected and re-digitized by the first author (RRD). Overall error was assessed using Intraclass Correlation Coefficients (ICCs). A two-way random-effects model with absolute agreement was used for interobserver error testing; a two-way random-effects model with consistency was used for intraobserver error testing.

To assess variation in cranial morphology and how it differs by collection, a series of one-way multivariate analyses of variance (MANOVAs), two-way MANOVAs, and two-way multivariate analyses of covariance (MANCOVAs) were conducted. These models quantified the impact of collection, and other demographic variables (e.g., sex, population), on cranial morphology. A Pillai trace test statistic was used for each MANOVA and MANCOVA, as it provides more robust results in the case of violations of the homogeneity of variance-covariance assumption (Ateş et al. 2019).

Canonical Variate Analysis (CVA) was conducted on each MANOVA model to identify

linear combinations of variables that best distinguish collections. The resultant canonical variates were used to visualize collection separation. This combined MANOVA and CVA approach was iteratively applied, starting with a single independent variable (collection) and progressively adding sex, population, donor age, birth year, and birth state. This approach clarified which variables contribute most to the separation between collections and assessed whether the effects of collection on cranial morphology remained significant after accounting for any variation in cranial morphology caused by other demographic variables.

Mahalanobis distances (D^2) were calculated to assess similarity/dissimilarity between the various collections and to explore potential hierarchical relationships. Mahalanobis distances were calculated for: 1) all data by collection; 2) European American data by collection; and, 3) all data by population, sex, and collection. Distance matrices were also calculated for the geographic distances between physical collection locations. To facilitate visual comparison between cranial morphology and geographic distances among collections, Mahalanobis distances and geographic distances were transformed using multidimensional scaling to create comparable coordinate spaces. The transformed datasets were then aligned using Procrustes analysis, allowing both sets of data to be plotted together for direct visualization of their spatial relationships.

To quantify the effect of collection-specific sample bias on cranial morphology, two linear discriminant functions were generated: 1) classification by collection; and, 2) classification into population and sex-specific groups. These functions assess whether classification rates are higher based on collection origin or traditional biological parameters (i.e., population and sex). Both models were cross-validated using a leave-one-out cross validation (LOOCV) procedure.

RESULTS

Initial Data Analysis

Demographic variables varied by collection (e.g., height in inches versus centimeters) so data cleaning was necessary, primarily related to simple formatting changes (e.g., capitalization, spacing, transforming state names to state codes). However, the language used to record population/ancestry information varied considerably by collection—some provided refined, population-level labels (e.g., Irish, Japanese, Mexican, etc.) others used general terms (e.g., Asian, European, Hispanic). Consequently, the more refined population labels were converted to general terminology to ensure comparability. We understand this limits some of the conclusions drawn from this study, but we were ultimately constrained by the information made available.

Sex, population, and age information was available for all individuals across all collections (Tables 1.1 and 1.3). Information on height, birth year, year of death, birth state, state at time of death, SES, and donation type varied by collection (see Table 1.3). STAFS DSC, TXSTDSC, and UTK DSC provided childhood SES, while JAWDHSC, TXSTDSC, and WMed STARS provided adulthood SES.

Table 1.3. Summary of mean statistics for collections on all factors assessed via ANOVAs, apart from population and sex which are provided in Table 1.

Collection	Age	Height (inches, M/F)	Birth Year	Year of Death	Top 3 Birth States	Top 3 States at time of Death	Top Reported SES*	Top Donation Type
JAWDHSC	66.3	69.6/64	1950	2017	NC, IL, PA	NC, VA, TN	Middle ^A	NOK
MMDSC	67.8	69.6/63.3	1928	1997	NM, NY, CA	NM, CA, MT	—	NOK
MLOC	78.1	—	—	—	—	—	—	—
MSUFAL DSC	53.4	69.3/64.6	1958	2012	MI, OH, AR	MI	—	NOK
STAFS DSC	66.6	69.5/64.5	1948	2015	TX, NY, IL	TX, MO, OK	Middle ^C	NOK
TXSTDSC	70.5	69.6/64	1945	2016	TX, CA, IN	TX, CA, TN	Middle ^{A,C}	NOK
UTK DSC	66.6	69.6/64.1	1945	2012	TN, KY, PA	TN, NC, GA	Middle ^C	Self
WMed STARS	71.3	68.6/64	1946	2018	MI, AR, IL	MI	Middle ^A	Self

*Childhood versus adulthood SES indicated by a “C” or “A”.

Due to antemortem tooth loss and alveolar resorption, the interlandmark distances with the most missing data were generally associated with measurements that include prosthion, such

as BPL, NPH, MAL, and UFHT (Figure 1.1). Following imputation, variable means were calculated for each collection (Table 1.4) to assess overall fit of the imputation process.

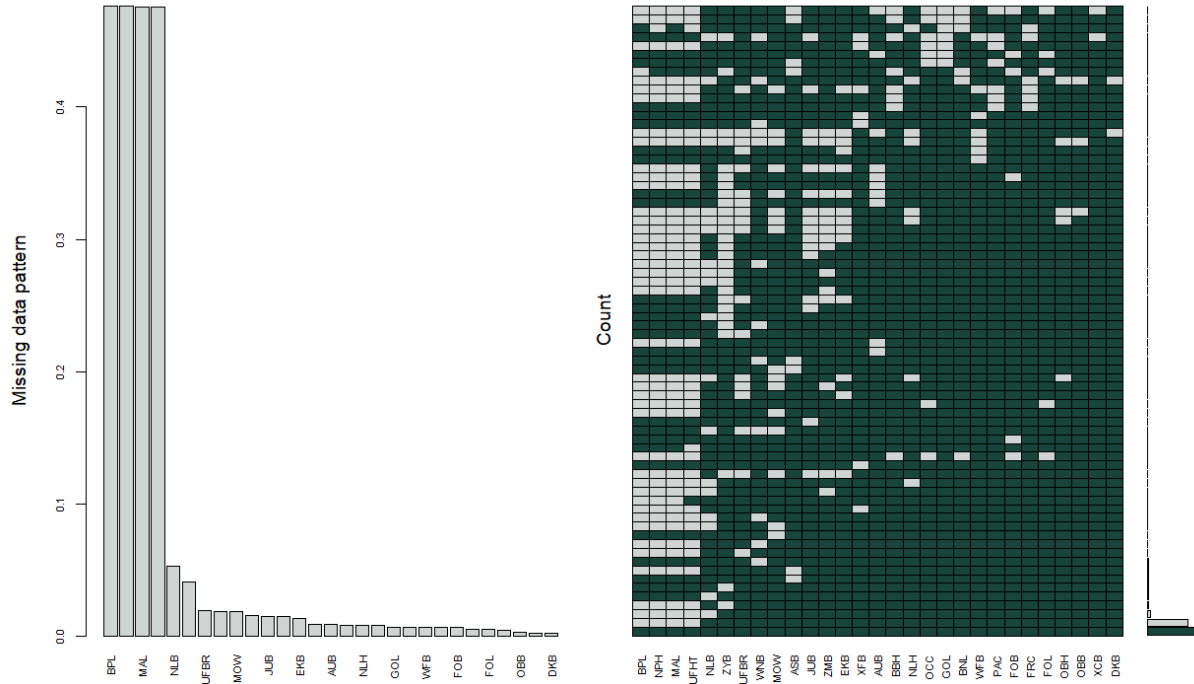


Figure 1.1. Visualization of missing data patterns among interlandmark distances across all collections. Missing values (grey) and observed values (green) are provided for each variable.

Table 1.4. Measurement means for each collection (pooled sexes) following imputation.

<i>Collections</i>	GOL			BNL			BBH			XCB			XFB		
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
JAWDHSC	184	7.8	34	103	5.1	27	137	6.2	32	138	5.5	27	117	5.5	24
MLOC	182	8.4	40	103	5.3	26	138	5.4	29	141	6.2	33	118	5.3	25
MMDSC	182	8.1	54	102	6.0	32	136	6.6	31	138	5.6	30	117	5.9	32
MSUFAL DSC	183	9.5	33	103	6.0	23	137	7.7	32	136	5.5	22	114	6.9	25
STAFS DSC	183	8.6	53	104	5.8	34	138	6.4	34	138	5.9	35	117	5.4	32
TXSTDSC	184	8.3	45	103	5.6	29	138	6.2	35	139	5.5	32	117	5.9	33
UTK DSC	183	8.5	46	102	5.2	25	136	6.1	30	139	5.7	33	117	5.8	36
WMed STARS	185	7.6	27	104	4.8	20	139	6.2	23	138	6.0	25	117	5.7	24
	WFB			ZYB			AUB			ASB			BPL		
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
JAWDHSC	95	4.9	25	124	5.8	29	120	5.2	26	112	4.8	29	95	5.7	28
MLOC	95	5.3	25	128	6.6	33	124	6.2	32	114	4.7	24	97	6.2	36
MMDSC	95	5.0	27	125	6.7	42	120	5.9	38	113	5.5	29	95	6.2	36
MSUFAL DSC	94	6.1	25	125	5.3	23	119	5.0	18	112	6.1	24	98	7.2	31
STAFS DSC	95	4.7	22	126	6.1	29	122	5.1	27	113	5.6	34	97	6.4	35
TXSTDSC	95	4.8	24	125	6.1	30	121	5.4	29	114	5.4	30	96	5.9	35
UTK DSC	95	4.9	33	125	5.9	29	122	5.4	30	113	5.4	34	95	5.9	29
WMed STARS	96	5.1	24	125	6.3	27	121	5.9	31	114	5.8	26	97	5.5	24
	NPH			NLH			JUB			NLB			MAL		
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
JAWDHSC	68	4.8	26	53	3.5	16	108	5.4	26	23	2.1	11	54	3.7	18
MLOC	69	3.6	19	54	3.3	16	112	6.0	28	25	2.3	12	54	3.6	20
MMDSC	69	4.3	25	54	3.6	19	109	5.8	34	24	2.2	13	54	4.0	28
MSUFAL DSC	69	3.9	19	53	3.2	16	108	5.0	23	24	2.2	10	55	4.4	20
STAFS DSC	69	4.1	24	54	3.4	17	110	5.0	30	24	2.1	13	54	3.8	20
TXSTDSC	69	4.2	24	53	3.5	20	109	5.1	25	24	1.9	12	54	3.6	20
UTK DSC	68	4.1	23	53	3.5	18	108	5.2	28	24	2.2	10	53	3.8	25
WMed STARS	69	3.2	15	54	2.5	11	109	4.9	20	24	2.0	8	55	3.0	12

Table 1.4. (cont'd)

	OBH			OBB			DKB			WNB			ZMB		
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
JAWDHSC	34	1.9	8	41	2.0	10	20	2.2	11	8	1.6	8	88	5.2	29
MLOC	35	2.1	12	41	2.0	8	21	2.2	11	8	2.0	10	92	6.4	30
MMDSC	34	1.9	11	40	2.1	10	21	2.2	12	8	1.8	11	89	5.3	29
MSUFAL DSC	34	1.8	8	41	2.4	13	20	2.5	13	8	1.7	8	88	5.9	24
STAFS DSC	34	2.0	10	41	1.8	11	21	2.0	12	9	1.6	11	89	4.9	28
TXSTDSC	34	2.0	10	41	2.0	11	20	2.1	13	9	1.8	11	89	5.0	30
UTK DSC	34	2.0	10	41	2.0	10	20	2.1	12	8	1.7	10	88	4.9	25
WMed STARS	34	2.0	10	41	1.9	9	21	2.1	8	9	1.8	8	88	4.2	17
	EKB			FRC			PAC			OCC			FOL		
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
JAWDHSC	96	4.2	19	112	5.0	28	116	6.8	34	99	5.9	32	37	2.7	18
MLOC	98	4.7	21	113	4.9	28	114	6.8	36	98	5.4	26	37	2.5	13
MMDSC	96	4.5	22	111	5.6	34	115	6.5	39	97	5.4	36	37	2.6	14
MSUFAL DSC	96	4.2	19	112	5.7	24	115	6.7	28	99	6.9	35	37	2.7	12
STAFS DSC	97	3.6	24	113	5.0	24	115	7.5	36	99	6.3	36	37	3.0	18
TXSTDSC	96	3.9	19	113	5.4	31	116	6.8	35	98	5.3	32	37	2.5	14
UTK DSC	96	4.2	25	112	5.4	28	115	6.6	40	98	5.5	37	37	2.7	13
WMed STARS	97	3.8	18	112	5.5	22	117	5.8	23	99	5.1	19	37	2.6	10
	FOB			MOW			UFBR			UFHT					
	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>			
JAWDHSC	32	2.7	13	52	5.6	29	102	4.8	25	70	5.0	27			
MLOC	31	2.3	11	53	5.9	30	104	4.7	20	71	3.7	20			
MMDSC	32	2.2	13	53	5.7	42	102	4.8	23	71	4.4	25			
MSUFAL DSC	31	2.5	10	52	5.8	23	102	4.5	20	71	3.9	18			
STAFS DSC	32	2.3	12	53	5.1	31	103	3.9	27	71	4.4	25			
TXSTDSC	32	2.4	13	52	5.2	28	103	4.6	26	71	4.4	26			
UTK DSC	32	2.6	13	52	5.5	44	102	4.6	26	71	4.2	24			
WMed STARS	32	2.3	12	52	5.3	25	103	4.3	21	71	3.3	15			

A two-way random-effects model for absolute agreement was used to assess interobserver error. The ICC was 0.999 (95% CI: 0.999 to 0.999), indicating excellent agreement. Using a two-way random-effects model for consistency, the intraobserver error test produced an ICC = 0.999 (CI: 0.999 to 1), indicating excellent consistency.

Analysis of Demographic Variables

Results of the one-way ANOVAs are provided in Table 1.5. Overall, sex ratios are relatively similar between collections. Although European Americans make up the largest population group in each collection (56-99%), population compositions varied significantly by collection, attributable to higher population diversity in the MLOC and lower diversity in JAWDHSC and UTK DSC (see the pairwise comparisons in Table 1.5). Mean adult age differed significantly between collections; again, these differences are attributed to a higher mean adult age in the MLOC and lower mean adult age in the MSUFAL DSC (see Table 1.5). Females were younger than males across all collections ($F(1,1156) = 17.3, p < 0.01$), but this is most prevalent at the MMDSC (adjusted $p < 0.01$).

Table 1.5. Results of one-way ANOVAs and post hoc Tukey tests on collection samples.

Factor of Comparison	F-value	DF	P-Value	Tukey Pairwise Comparisons*	Adjusted P-Value
Sex	2.05	7, 1153	0.047	All comparisons	>0.05
Population	6.15	7, 1153	<0.01	MLOC v. JAWDHSC, MMDSC, and UTK DSC	<0.01
				MSUFAL DSC v. JAWDHSC	<0.05
				TXSTDSC v. JAWDHSC	<0.01
				TXSTDSC v. UTK DSC	<0.05
Age	14.66	7, 1150	<0.01	MLOC v. JAWDHSC, MMDSC, MSUFAL DSC, STAFS DSC, and TXSTDSC	<0.01
				MSUFAL DSC v. JAWDHSC, MMDSC, STAFS DSC, TXSTDSC, UTK DSC, and WMed STARS	<0.01
Height	0.75	6, 917	>0.05	—	—
Birth Year	57.1	6, 1045	<0.01	JAWDHSC v. TXSTDSC	<0.01
				JAWDHSC v. UTK DSC	<0.05
				MMDSC v. JAWDHSC, MSUFAL DSC, STAFS DSC, TXSTDSC, UTK DSC, and WMed STARS	<0.01
				MSUFAL DSC v. STAFS DSC, TXSTDSC, UTK DSC, and WMed STARS	<0.01
Year of Death	168.7	6, 1050	<0.01	MMDSC v. JAWDHSC, MSUFAL DSC, STAFS DSC, TXSTDSC, UTK DSC, and WMed STARS	<0.01
				MSUFAL DSC v. JAWDHSC and WMed STARS	<0.05
				UTK DSC v. JAWDHSC, TXSTDSC, and WMed STARS	<0.01
				UTK DSC v. STAFS DSC	<0.05
Birth State	9.85	6, 795	<0.01	MMDSC v. STAFS DSC, TXSTDSC, and UTK DSC	<0.01
				MSUFAL v. STAFS DSC	<0.01
				WMed STARS v. JAWDHSC, STAFS DSC, TXSTDSC, and UTK DSC	<0.01
State at Time of Death	164.4	6, 727	<0.01	**	<0.01
Childhood SES	0.48	2, 595	>0.05	—	—
Adulthood SES	1.09	2, 313	>0.05	—	—

*Only significant pairwise comparisons included for significant ANOVAs.

**All comparisons were significant at $p < 0.01$ apart from three comparisons: JAWDHSC v. UTK DSC ($p > 0.05$), MSUFAL v. WMed STARS ($p > 0.05$), and STAFS DSC v. TXSTDSC ($p > 0.05$).

Mean height did not differ significantly by collection or birth year. Although all of the collections are modern, significant differences in mean birth year and year of death were identified (see Table 1.5). The MMDSC in New Mexico has a significantly earlier mean birth year than all other collections (Figure 1.2). Conversely, the MSUFAL DSC has a much later mean birth year than all others except the JAWDHSC (see Table 1.5). The MMDSC had significantly earlier mean year of death than all other collections. The UTK DSC also had significantly earlier mean year of death than JAWDHSC, STAFS DSC, TXSTDSC, and WMed STARS (see Table 1.5).

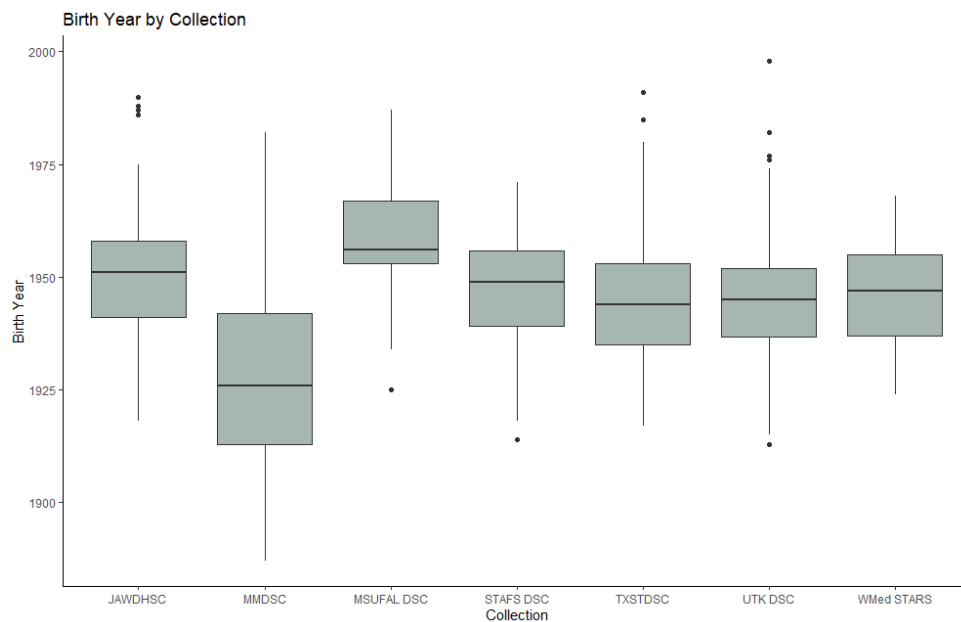


Figure 1.2. Boxplot of birth years, by collection.

Significant differences were identified between collections for state of birth and state at time of death. These are visualized in Figure 1.3 and further outlined in Table 1.5. Perhaps unsurprisingly, individuals in a collection are most often from the same state where the collection is located. There was no significant difference in childhood or adult SES between collections, as most individuals reported a middle SES (see Table 1.3).



Figure 1.3. Proportional heat map for states of birth (A) and states at time of death (B) for donors in each collection.

Craniometric Analysis

Cook's distance identified six potential outliers. After direct observation of these data and associated boxplots and pairwise plots, one European American female from the MSUFAL DSC was removed from the dataset. Upon reexamination of the skeletal remains, the outlier status of this individual (i.e., small cranial dimensions) was directly associated with a pathology.

Individual results of the MANOVA and MANCOVA analyses are provided in Table 1.6. Collection origin consistently had a significant effect on cranial morphology, even after accounting for variation in cranial morphology caused by other demographic variables (see Table 1.6). CVA and canonical variate visualizations demonstrate how collection separation varies based on the demographic variables included (Figure 1.4). With collection as the sole independent factor in the MANOVA and CVA, the second canonical variate separates the MMDSC and MLOC from the other six collections (Figure 1.4A). The separation between MLOC and the other collections decreases after sex and population are added to the model (Figure 1.4B). However, across all models that incorporate collection and additional demographic variables, the first canonical variate separates the MMDSC from all other collections (Figures 1.4B-D).

Table 1.6. Multivariate MANOVA and MANCOVA summaries.

Analysis	Factor	DF	Pillai	Approx. F	num df	Error df	Pr (>F)
One-Way MANOVA	Collection	7	0.49	2.95	203	7,910	<0.01
Two-Way MANOVA	Collection	7	0.50	3.01	203	7,903	<0.01
	Sex	1	0.62	63.0	29	1,123	<0.01
Two-Way MANOVA	Collection	7	0.53	3.14	203	7,819	<0.01
	Sex	1	0.62	63.2	29	1,111	<0.01
	Population	12	0.66	2.24	348	13,464	<0.01
Two-Way MANCOVA	Collection	7	0.54	3.19	203	7,798	<0.01
	Sex	1	0.62	63.0	29	1,108	<0.01
	Population	12	0.66	2.25	348	13,428	<0.01
	Age	1	0.12	5.05	29	1,108	<0.01
Two-Way MANCOVA	Collection	6	0.43	2.66	174	6,060	<0.01
	Sex	1	0.64	61.8	29	1,005	<0.01
	Population	8	0.48	2.23	232	8,096	<0.01
	Age	1	0.12	4.94	29	1,005	<0.01
	Birth Year	1	0.09	3.41	29	1,005	<0.01
Two-Way MANCOVA	Collection	6	0.46	2.67	174	5,622	<0.01
	Sex	1	0.66	61.9	29	932	<0.01
	Population	8	0.52	2.27	232	7,512	<0.01
	Age	1	0.13	4.96	29	932	<0.01
	Birth Year	1	0.10	3.36	29	932	<0.01
	Birthplace	73	2.23	1.10	2117	27,840	<0.01
Two-Way MANCOVA (Reduced MMDSC Data)	Collection	6	0.47	2.58	174	5,316	<0.01
	Sex	1	0.66	58.4	29	881	<0.01
	Population	8	0.54	2.20	232	7,104	<0.01
	Age	1	0.14	5.06	29	881	<0.01
	Birth Year	1	0.10	3.49	29	881	<0.01
	Birthplace	73	2.35	1.10	2117	26,361	<0.01

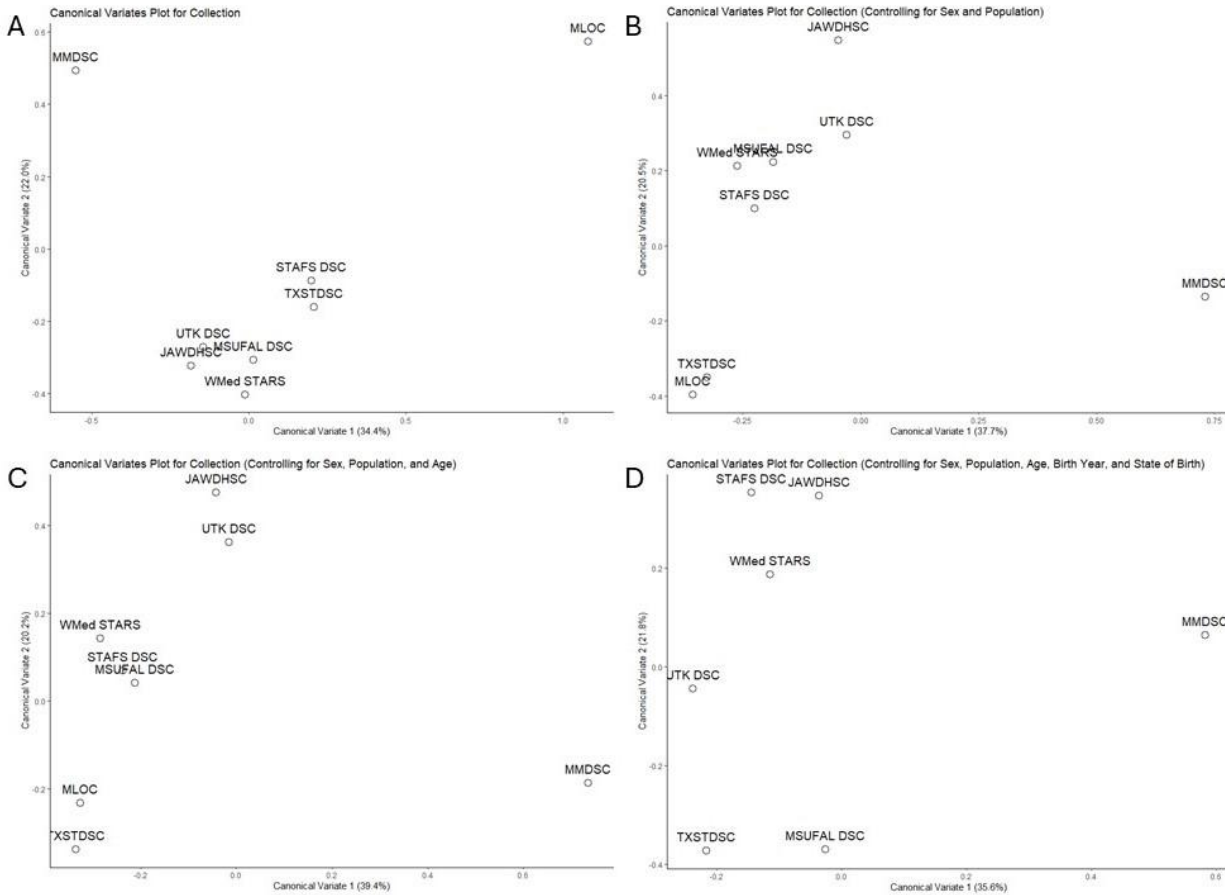


Figure 1.4. Canonical Variates Plots of interlandmark distances by: A) collection; B) collection, accounting for variation caused by sex and population; C) collection, accounting for variation caused by sex, population, and age; and, D) collection, accounting for variation caused by sex, population, age, birth year, and state of birth (MLOC not included as birth year or state of birth information was not available). All plots display separation of collections using the first two canonical variates (Can1 and Can2), which accounted for 52.4-59.6% of the variations, depending on the model used.

The questionable identification status of the individuals from the MMDSC (Komar and Grivas 2008) required a MANCOVA with a reduced MMDSC dataset to determine if these individuals were driving the separation of MMDSC from the other seven collections (see Table 1.6). All of the “transfer” designated individuals from the Department of Anatomy or the Office of the Medical Examiner, and those without associated donation documentation, were removed from the analysis (Table 1.7). The CVA and visualization of collection separation using canonical variates 1 and 2 (Figure 1.5) further separates the MMDSC from the other six collections,

suggesting identification status was not contributing to this separation. Across all MANCOVAs, including analyses with both MMDSC datasets, the most influential factor on cranial variation is sex, followed by population, collection, state of birth, age, and birth year.

Table 1.7. Breakdown of reduced MMDSC sample by population and sex.

Sample (by Collection and Population)	Male (<i>n</i>)	Female (<i>n</i>)
Maxwell Museum's Documented Skeletal Collection (<i>n</i> = 187)		
African American	4	2
European American	93	80
Hispanic	6	2
Multiple	0	0
Total	103	84

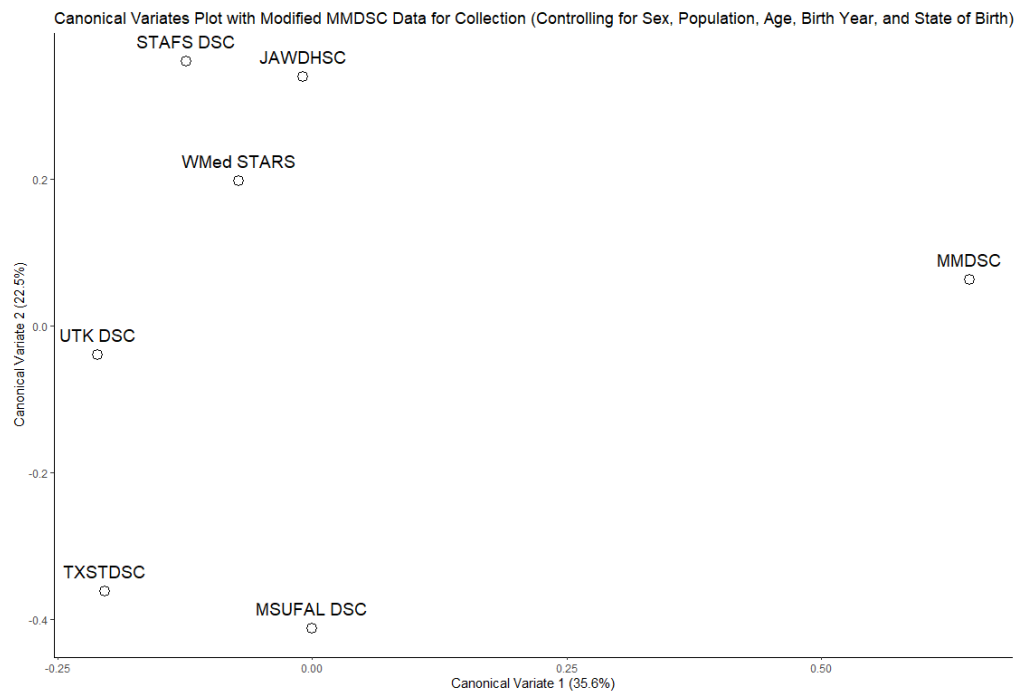


Figure 1.5. Canonical Variates Plot of interlandmark distances with the reduced MMDSC data by collection, accounting for variation caused by sex, population, age, birth year, and state of birth (MLOC not included as birth year or state of birth information was not available). The plot displays the first two canonical variates (Can1 and Can2), which account for 35.5% and 22.4% of the variance, respectively. MMDSC shows the greatest divergence along Can1.

Mahalanobis distance (D^2) matrices were calculated and are illustrated in Figures 1.6-1.7. A Mahalanobis distance (D^2) matrix between collections, populations, and sex-specific groups is presented in Figures 1.8-1.9.

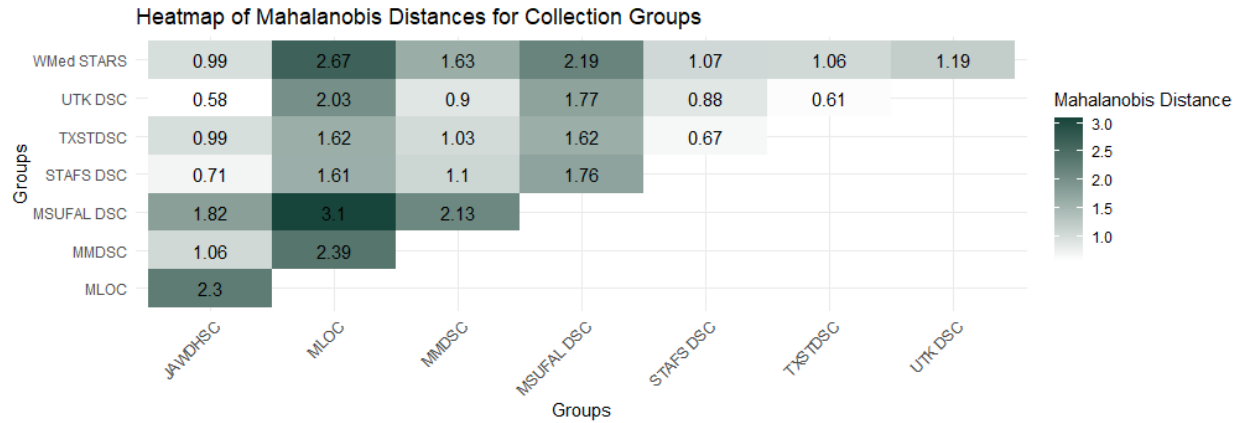


Figure 1.6. Heatmap of the Mahalanobis distance (D^2) matrix calculated based on collection origin. Greater distances are represented by darker green tones.

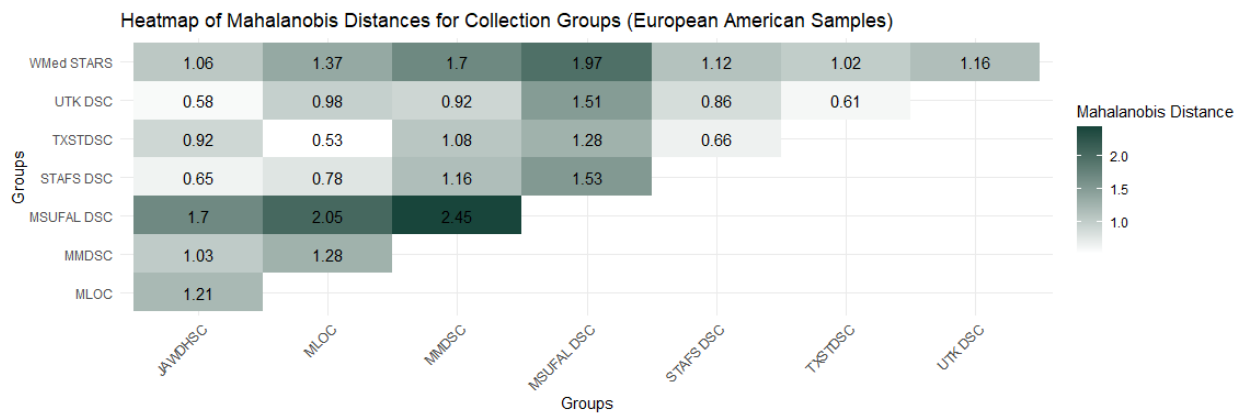


Figure 1.7. Heatmap of the Mahalanobis distance (D^2) matrix calculated based on collection origin for European American samples. Greater distances are represented by darker green tones.

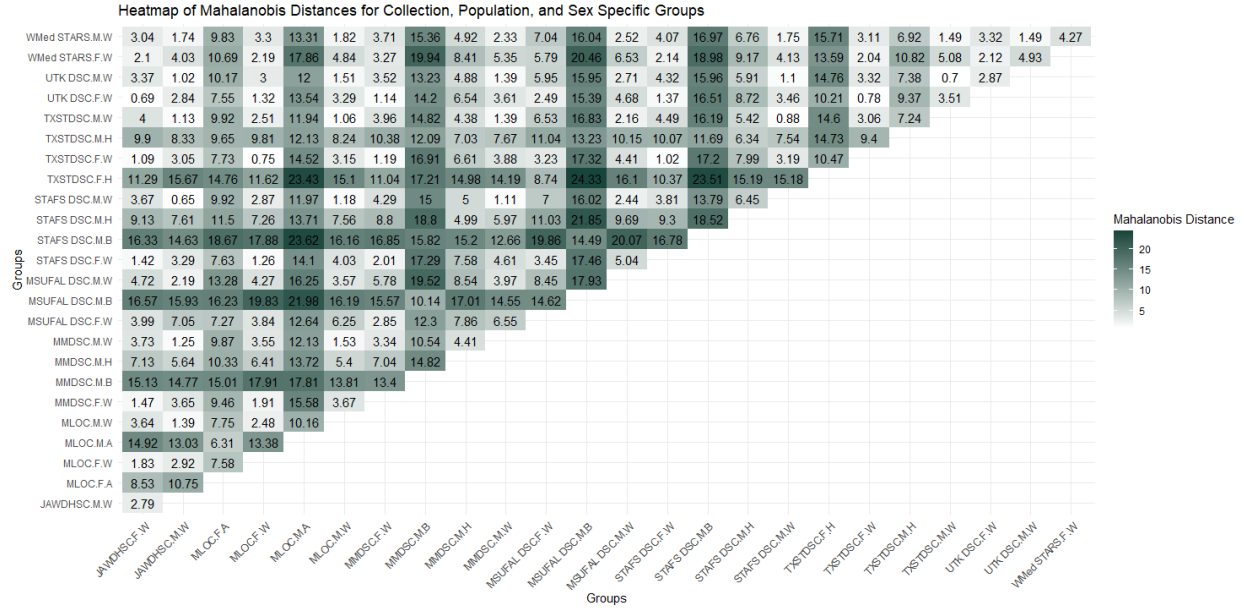


Figure 1.8. Heatmap of the Mahalanobis distance (D^2) matrix calculated based on collection, population, and sex specific groups. “M” and “F” are code for male and female. “A”, “B”, “H”, and “W” are code for Asian, African American, Hispanic, and European American, respectively. Greater distances are represented by darker green tones.

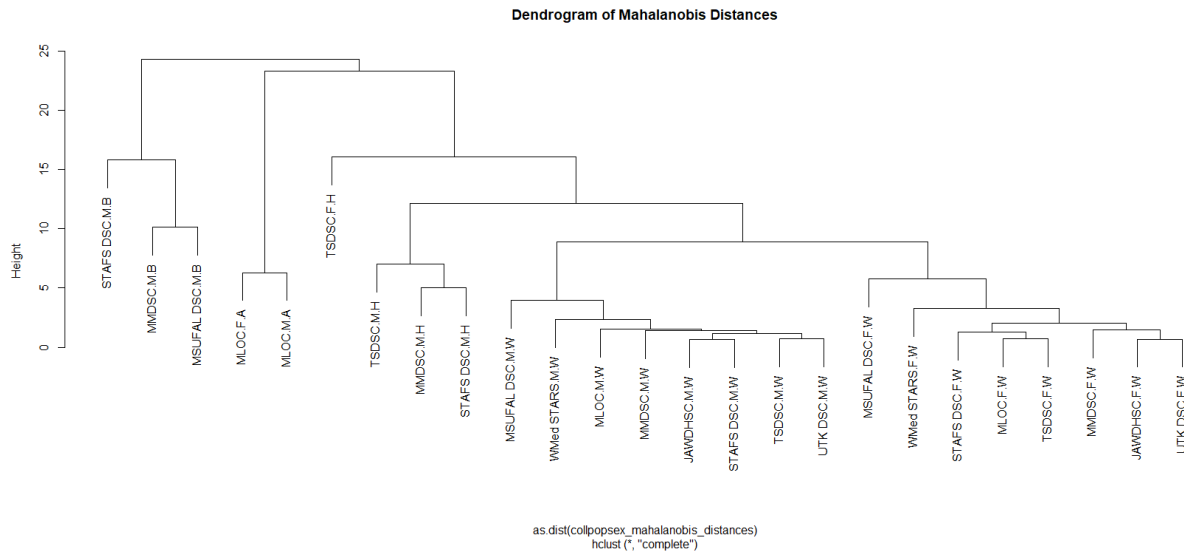


Figure 1.9. Hierarchical clustering dendrogram based on the Mahalanobis distance (D^2) matrix calculated based on collection, population, and sex specific groups. “M” and “F” are code for male and female. “A”, “B”, “H”, and “W” are code for Asian, African American, Hispanic, and European American, respectively.

Geographic distance between collection locales is illustrated in Figure 1.10. Plotting collection craniometric data and geographic locations together shows that the MSUFAL DSC and

WMed STARS have the greatest separation between craniometric data and geographic location (Figure 1.11). Plotting craniometric and geographic data of the European American samples controls for any population differences driving craniometric separation and highlights the relationship between collections and geography. In Figure 1.12, the MLOC, MMDSC, MSUFAL DSC, and WMed STARS show unexpected differences from their geographic location; while the MLOC craniometric sample moves closer to the other European craniometric samples, the MMDSC moves farther away.

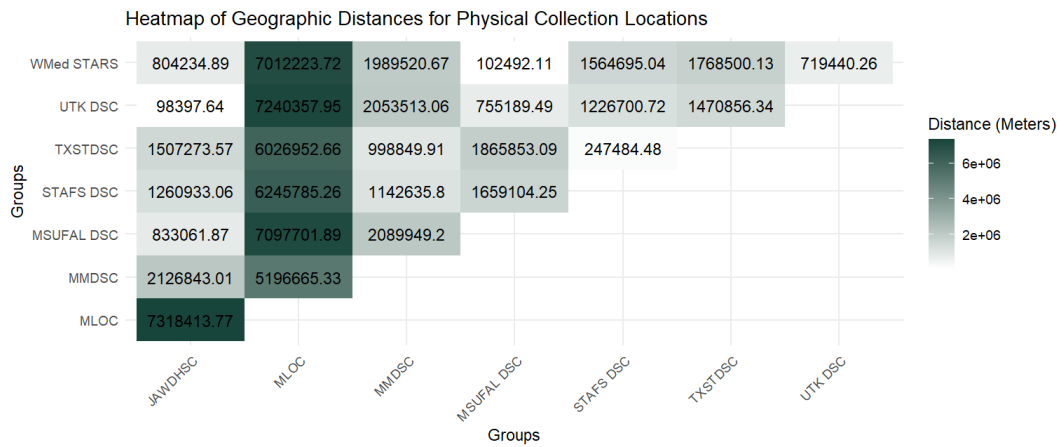


Figure 1.10. Heatmap of the distance matrix calculated based on physical collection locations (i.e., GPS coordinates). Distances are in meters and greater distances are represented by darker green tones.

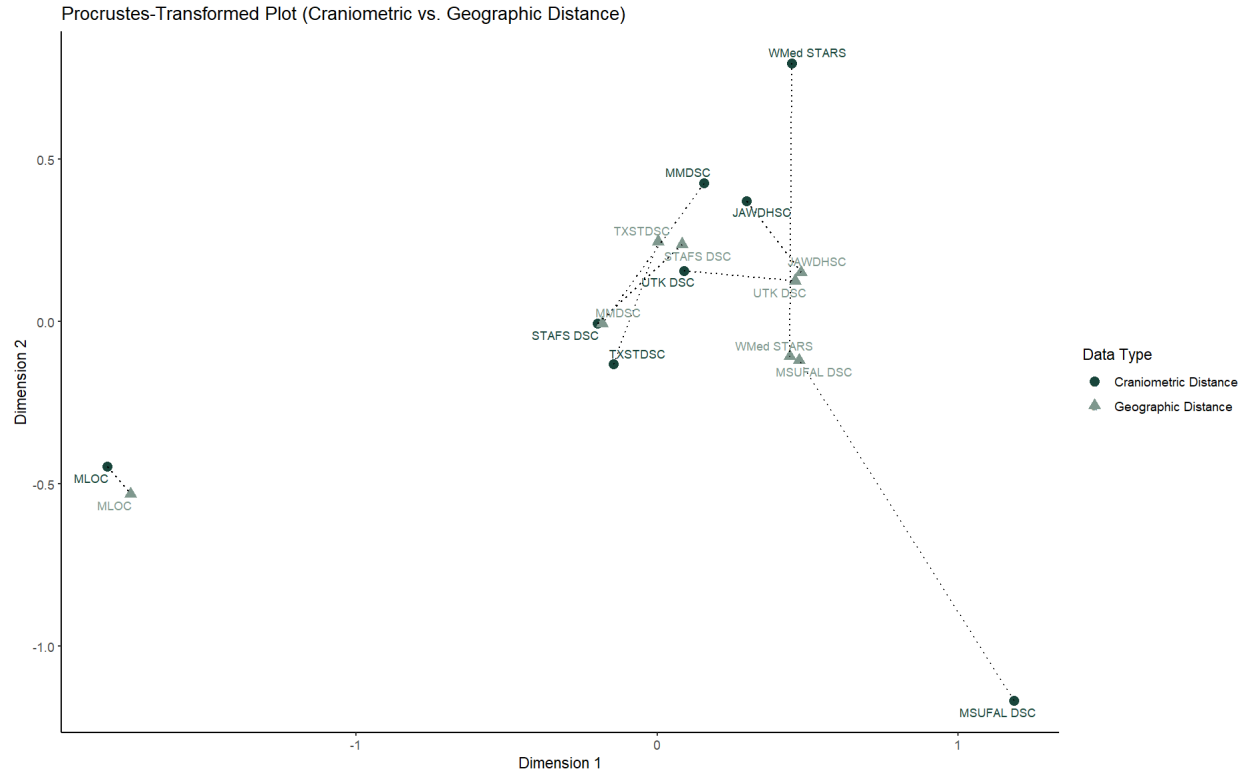


Figure 1.11. Procrustes transformed data of cranial morphology and geographic distance among collections. Mahalanobis distances based on cranial measurements (dark green circles) and geographic distances based on collection locations (light green triangles) were transformed using multidimensional scaling and aligned via Procrustes analysis. Dotted lines connect craniometric centroids and geographic location for each collection.

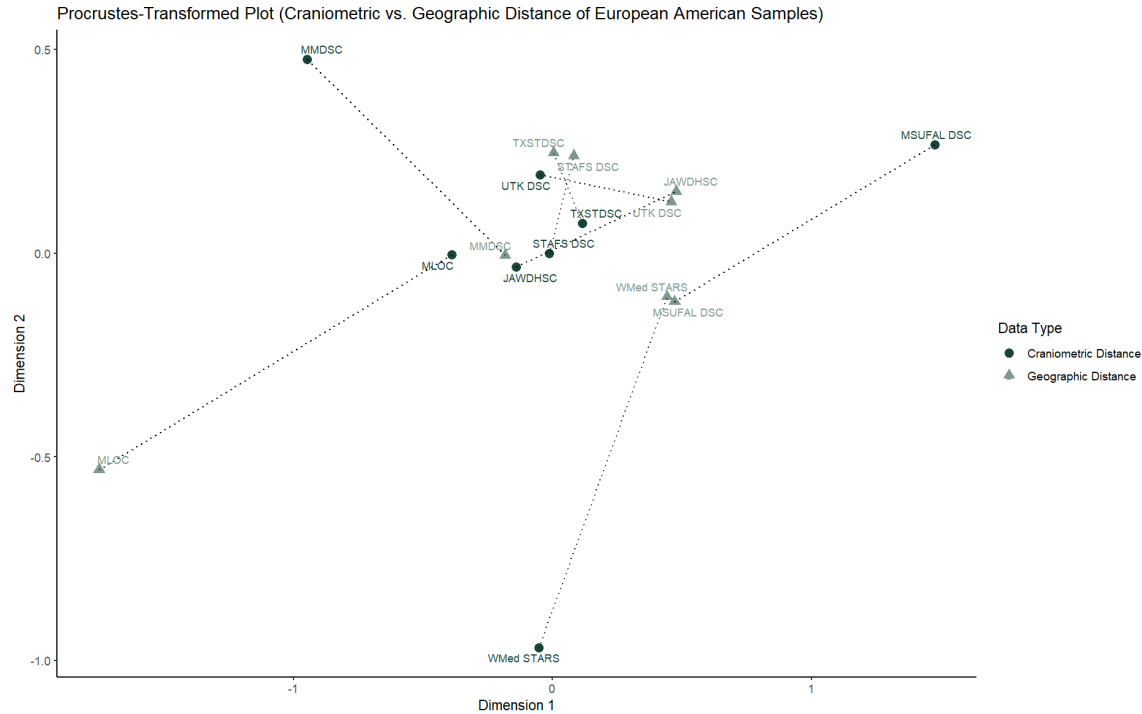


Figure 1.12. Procrustes transformed data of cranial morphology and geographic distance among collections using only European American samples. Mahalanobis distances based on cranial measurements (dark green circles) and geographic distances based on collection locations (light green triangles) were transformed using multidimensional scaling and aligned via Procrustes analysis. Dotted lines connect craniometric centroids and geographic location for each collection.

A linear discriminant function was used to classify all individuals by collection. Due to small sample sizes, the MSUFAL DSC and WMed STARS were excluded. The LOOCV model achieved an overall performance of 32.0% (1.92 times better than chance), with individual group classification rates ranging from 7.6% for JAWDHSC to 45.5% for MMDSC (Table 1.8).

Table 1.8. Classification rate by collection origin (LOOCV).

<i>Reference</i>	<i>Prediction</i>						Total	CCR (%)
	JAWDHSC	MLOC	MMDSC	STAFS DSC	TXSTDSC	UTK DSC		
JAWDHSC	9	4	24	15	32	34	118	7.6%
MLOC	2	32	17	8	31	10	100	32.0%
MMDSC	9	13	112	14	53	45	246	45.5%
STAFS DSC	10	11	33	21	53	22	150	14.0%
TXSTDSC	7	14	55	20	109	45	250	43.6%
UTK DSC	10	9	55	18	64	64	220	29.1%
						Overall	1084	32.0%

A second linear discriminant function was created to classify individuals by population and sex; to balance samples, small population and sex groups were removed and larger groups downsampled. Overall performance of this model was 49.0% (or, 3.92 times better than chance), with individual group classification rates ranging from 33.3% for Asian males and females to 66.7% for European males (Table 1.9).

Table 1.9. Classification rate of individuals by population and sex (LOOCV). “M” and “F” are code for male and female.

<i>Reference</i>	<i>Prediction</i>								Total	CCR (%)
	African American F	African American M	Asian F	Asian M	European American F	European American M	Hispanic F	Hispanic M		
African Amer F	6	3	0	0	0	0	2	1	12	50.0%
African Amer M	3	7	0	1	0	0	1	0	12	58.3%
Asian F	0	1	5	1	4	0	0	1	12	41.7%
Asian M	0	1	4	4	1	2	0	0		33.3%
European Amer F	1	1	0	2	4	3	1	0	12	33.3%
European Amer M	1	0	0	0	0	8	0	3	12	66.7%
Hispanic F	1	0	2	1	1	0	7	0	12	58.3%
Hispanic M	0	1	1	1	1	2	0	6	12	50.0%
								Overall	96	49.0%

DISCUSSION

This paper addresses whether collection-specific sample bias is reflected in skeletal variation. First, demographic variables were compared across collections to reveal potential sources of bias among the various collections. Second, craniometric data from each collection served as a potential proxy to quantify how sample bias may influence biological data results and whether that bias obscures expected patterns of human variation. Together, these analyses provide a broad perspective on how modern documented skeletal collections samples impact our understanding of human variation and influence the methods developed using that material.

Demographic Sample Bias

We identified significant differences in demographic composition across all collections. And while no single collection was significantly different from all other collections, *all* of the collections *were* significantly different from others for at least one demographic variable (see Table 1.5). These analyses suggest that demographic variables associated with each collection—whether intentional, such as procurement strategies, or unintentional, like geographic location—likely shape the donor population and result in unique samples that are characteristic of the individual collections.

This analysis highlights pitfalls potentially encountered if only one documented skeletal collection is used for a study. For example, the MLOC may not provide the best sample for age estimation methods since it is predominately made up of significantly older individuals; however, if a researcher is interested in a diverse sample (ancestral and ethnicities), the MLOC may be one of the better choices. These sample biases are not detrimental, but they do have the potential to impact a study if researchers are unaware of potential biases. As Albanese (2018) said, to employ appropriate sampling strategies, the researcher is responsible for understanding the representativeness of a collection and any inherent sample bias therein.

Collection-Specific Cranial Variation

We were able to demonstrate that *collection origin* does influence the collective cranial morphology represented in modern collections ($p < 0.01$). However, collection is not the most influential factor of that variation—instead, sex has the greatest impact, followed by population, collection, state of birth, age, and birth year.

This supports the earlier work of Moore-Jansen (1989), who, after investigating collection-specific cranial variation between the historic Hamann-Todd and Terry Collections, found sex was more influential on cranial morphology than collection origin. And although he did not directly assess population variability, he did attribute the significant differences seen between European Americans in each collection to differences in European groups represented by each. During the early 20th century when both collections were actively acquiring cadavers, there were many European immigrants settling in the U.S.—Germans and British in and around St. Louis, MO, where the Terry Collection originated, and eastern and southern Europeans in Cleveland, OH, home to the Hamann-Todd Collection (Moore-Jansen 1989). As poor immigrants comprised a large amount of the European sample in those historic collections (Moore-Jansen 1989; Muller et al. 2017), the differences he noted very likely reflect population differences. However, Moore-Jansen (1989) argues modern collections are not impacted by secular change and have fewer immigrant samples, and thus will avoid collection-specific sample biases—a finding disputed by our study.

Visualizing patterns in these data to highlight some of the effect sample bias plays on cranial morphology suggested geographic clusters and separation. For example, the JAWDHSC, MSUFAL DSC, STAFS DSC, TXSTDSC, UTK DSC, and WMed STARS all pair with their closest geographic neighbor (see Figure 1.4A). However, after accounting for variation in cranial morphology caused by sex and population (see Figure 1.4B), changes in clusters and separation

indicate population composition at each collection may have a confounding effect. Population, more than anything else, may be contributing to these “geographic” clusters, since collections from similar geographic regions likely contain similar donor demographic profiles (i.e., donors from similar population groups).

Further investigation using plots of both craniometric and geographic distances reinforces this argument (see Figures 1.11 and 1.12). By reducing the sample to only European Americans the effect of population on cranial morphology is removed. Now, the craniometric samples no longer cluster near collection location. The original “geographic” clusters represent higher proportions of African Americans (MSUFAL DSC), Asian Americans (MLOC), Hispanic individuals (STAFS DSC and TXSTDSC), and European Americans (JAWDHSC, MMDSC, and UTK DSC). Geographical location is only correlated with variation in cranial morphology because collections from similar geographic regions have similar donor populations, resulting in samples with similar population compositions. There is no evidence that cranial morphology varies by geographic region for individuals within the same population group in the United States.

Divergence of the MSUFAL DSC and WMed STARS from the other collections may result from smaller sample sizes, but there is some confusion regarding the MMDSC sample and why it is separating from all other collections when controlling for population (see Figures 1.4B-D and 1.12). Neither sample size, secular change, nor location of birth explains this phenomenon. The potential influence of poor documentation was considered, so we removed individuals from the MMDSC sample who were poorly documented (see Figure 1.5). This increased the collection’s divergence, indicating those individuals were not driving the separation. Inter- and intraobserver error are not influencing the distinctness of the MMDSC either. Childhood SES could potentially play a factor, as Cardoso (2007) suggested low SES status can impact skeletal morphology.

Unfortunately, this information was not available for all collections, including the MMDSC, so it was not possible to test that hypothesis.

There may be more individuals included in our study sample who have an incorrectly documented social race or are otherwise insufficiently documented (e.g., multiple social race categories not recorded), particularly when we consider the poor documentation of the MMDSC during its earliest years (Komar and Grivas 2008). However, there are also a large number of Hispanic and New Mexican Hispanic individuals in New Mexico and, importantly for our consideration, the census definitions of Hispanic as a social race versus ethnicity has shifted over time (ACS 5-Year Estimate 2022; Komar and Grivas 2008). Documentation of donors reflects definitions and perceptions of social race at the time of donation (Albanese 2018); perhaps documentation was correctly recorded for the individuals within the MMDSC, but those individuals would fall under the guise of a different population today, confounding studies using population information. Ultimately, something is setting the MMDSC apart from other collections. This may reflect unique populations (i.e., New Mexican Hispanic) or the MMDSC is a more biased sample from poor or outdated documentation. If the former, future studies should compare the MMDSC to other collections with similar population histories. If the latter, the MMDSC may be less representative of the modern U.S. population.

Although collection origin does influence the collective cranial morphology represented at each collection, we demonstrate that it *does not* obscure our understanding of human skeletal variation. Samples will consistently and preferentially cluster with individuals from similar population and sex groups rather than with individuals from the same collection of origin (see Figures 1.8-1.9). This is *contra* Albanese (2018), who compared cranial indices among individuals in the Terry and Coimbra Collections and identified more similarities in cranial dimensions

between European and African Americans from the same collection (i.e., the Terry Collection) than between European Americans and Europeans in different collections (i.e., the Terry and Coimbra Collections). Using those data, he argues the groupings by peer-perceived population groups are a misinterpretation of sample bias and instead represent mortality bias and secular change unique to each collection. If Albanese's interpretation of those data is correct, then samples in our study should have primarily and consistently grouped by collection and not by other factors. As our samples did not cluster by collection despite their similar temporal period and geographic origin (i.e., all modern collections from the United States), the disparate temporal periods and geographic origins of the Terry and Coimbra Collections may explain some level of Albanese's (2003, 2018) results.

Classification accuracies further demonstrated that sample bias does not obscure human variation, as higher accuracies were achieved by classifying individuals by population and sex rather than by collection. If sample bias creates artificial "populations" that mask genuine patterns of human skeletal variation, classification by collection origin would be much higher. The collection classification model achieved a lower accuracy despite a much larger sample size and fewer classification categories, so it would appear differences in cranial morphology due to sample bias may be difficult to quantify. One limitation for this comparison is the small sample sizes used for the classification by population and sex; however, numerous studies have demonstrated individuals can be accurately classified into population and sex groups, at rates even higher than those achieved in this study (Dunn et al. 2020). These analyses verify that normal patterns of human variation are not obscured by collection origin and therefore documented skeletal collections are valid and reliable sources of data for understanding population diversity and sex differences, even when some level of collection-specific bias is present.

Implications for Biological and Forensic Anthropology

This study has two potential repercussions for biological and forensic anthropology. First, the significant differences identified among demographic variables in each collection suggest some level of collection-specific sample bias. Therefore, researchers need to understand the history and context of their study collection prior to data collection and analysis. Ethical considerations for body donation and curation have increased the number of peer-reviewed scientific papers on documented skeletal collections (e.g., Alves-Cardoso et al. 2022). However, information is not available for every collection and published information becomes outdated as quickly as new donors are added. Fortunately, curators and collection managers can provide invaluable information to visiting researchers—in fact, this study would not have been possible without their aid and insight during data collection. Although sample bias exists, understanding collection limitations and employing proper sampling procedures can account for many impacts of collection-specific sample bias (Albanese 2018).

Second, the analysis of craniometric data of individuals from eight modern documented skeletal collections within the same country of origin revealed that, while collection-specific sample bias influences cranial morphology, its effect does not obscure our current understanding of human skeletal variation. Other biological parameters, such as population and sex, still have a greater impact on cranial morphology, corresponding to previous assumptions in the field. These findings are important to the discipline, as the use of population and sex-specific methods are essential for the estimation of an accurate biological profile. Modern collections are therefore appropriate sources of skeletal samples for developing and validating analytical methods in biological and forensic anthropology, and previously published research findings are not merely misinterpretations of collection-specific sample bias.

CONCLUSION

This study investigated the existence of collection-specific sample bias and its impact on our understanding of normal human skeletal variation. We identified and quantified evidence of collection-specific sample bias, with significant demographic differences between collections. These findings suggest researchers must assess bias within a collection to employ appropriate sampling strategies effectively.

Our analysis also found that collection-specific sample bias does affect our understanding of variation in cranial morphology, although its influence is weaker than sex and/or population. Consistent with established assumptions in biological anthropology, individuals from the same population and sex groups exhibit more similar cranial morphology than those from the same documented collection. These results support the continued use of documented skeletal collections for method development and validation, refuting claims that population affinity studies are merely artifacts of sample bias patterns.

To further explore sample bias patterns, future studies should investigate collection-specific biases in the southwestern United States, conduct similar analyses with other skeletal traits (e.g., nonmetric traits), and assess the potential influence of childhood socioeconomic status, which was excluded here due to limited records.

BIBLIOGRAPHY

- Albanese, J. 2003. *Identified skeletal reference collections and the study of human variation*. (Unpublished doctoral dissertation. Ottawa, Canada: McMaster University.
- Albanese, J. 2018. Strategies for dealing with bias in identified reference collections and implications for research in the 21st century. In: C. Y. Henderson & F. A. Cardoso, eds. *Identified skeletal collections: The testing ground of anthropology?* Oxford, UK: Archaeopress Publishing LTD. pp. 59-82.
- Alves-Cardoso, F., Campanacho, V., & Plens, C. R. 2022. Topical collection “The Rise of Forensic Anthropology and Documented Human Osteological Collections”. *Forensic Sciences* 2(3):551-555. <https://doi.org/10.3390/forensicsci2030039>
- Armelli, K., Dunn, R. R., & Isaac, C. V. 2022. Understanding skeletal and body donation in forensic anthropology: Developments and procedures. Paper presented at the 74th Annual Conference of the American Academy of Forensic Sciences in Seattle, Washington.
- Ateş, C., Kaymaz, Ö., Kale, H., & Tekindal, M. A. 2019. Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of Type-I error rates. *Computation and Mathematical Methods* 2019(1):1-8.
- Avelar, T., Eduardo, L., Cardoso, M. A., Santos Bordoni, L., de Miranda Avelar, L., & de Miranda Avelar, J. V. 2017. Aging and sexual differences of the human skull. *Plastic and Reconstructive Surgery – Global Open* 5(4): e1297. DOI: 10.1097/GOX.0000000000001297
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. 2011. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research* 20(1): 40-49. DOI: 10.1002/mpr.329
- Bass, B., & Jefferson, J. 2003. *Death's acre: Inside the legendary forensic lab The Body Farm where the dead do tell tales*. New York, NY: C. P. Putnam's Sons.
- Body Donation. 2024. Forensic Anthropology Center, University of Tennessee, Knoxville. Retrieved October 13, 2024, from <https://fac.utk.edu/body-donation/>
- Campanacho, V., Ales Cardoso, F., & Ubelaker, D. H. 2021. Documented skeletal collections and their importance in forensic anthropology in the United States. *Forensic Sciences* 1:228-239. <https://doi.org/10.3390/forensicsci1030021>
- Cardoso, H. F. V. 2007. Environmental effects on skeletal versus dental development: Using a documented subadult skeletal sample to test a basic assumption in human osteological research. *American Journal of Biological Anthropology* 132:223-233.
- Cohen, M. N., Wood, J. W., & Milner, G. R. 1994. The osteological paradox reconsidered and reply. *Current Anthropology* 35(5): 629-637.

- de la Cova, C. 2019. Marginalized bodies and the construction of the Robert J. Terry Collection: A promised land lost. In: M. L. Mant & A. J. Holland, eds. *Bioarchaeology of Marginalized People*. London, UK: Springer Inc. pp. 133-155. <https://doi.org/10.1016/C2017-0-02300-5>
- de la Cova, C. 2021. Making silenced voices speak: Restoring neglected and ignored identities in anatomical collections. In: CM Cheverko, JR Prince-Buitenhuys, & M Hubbe, eds. *Theoretical approaches in bioarchaeology*. New York, NY: Routledge Taylor & Francis Group. pp. 150-169. <https://doi.org/10.4324/9780429262340>
- Dunn, R. R., Spiros, M. C., Kamnikar, K. R., Plemons, A. M., & Hefner, J. T. 2020. Ancestry estimation in forensic anthropology: A review. *WIREs Forensic Science* 2:e1369. <https://doi.org/10.1002/wfs2.1369>
- Fleischman, J. M., & Crowder, C. M. 2019. Standard operation procedure for microscribe 3-dimensional digitizer and craniometric data.
- George, R. L., Zejdlik, K., Messer, D. L., & N. V. Passalacqua. 2022. The John A. Williams Human Skeletal Collection at Wester Carolina University. *Forensic Sciences* 2:362-370. <https://doi.org/10.3390/forensicsci2020026>
- Gocha, T. P., Mavroudas, S. R., & Wescott, D. J. 2022. The Texas State Donated Skeletal Collection at the Forensic Anthropology Center at Texas State. *Forensic Sciences* 2(1):7-19. <https://doi.org/10.3390/forensicsci2010002>
- Gordon, C. C., & Bradtmiller, B. (1992). Interobserver error in a large scale anthropometric survey. *American Journal of Human Biology*, 4(2): 253-263. <https://doi.org/10.1002/ajhb.1310040210>
- Henderson, C. Y. 2018. Introduction. In: C. Y. Henderson & F. A. Cardoso, eds. *Identified skeletal collections: The testing ground of anthropology?* Oxford, UK: Archaeopress Publishing LTD. pp. 1-10.
- Jantz, L. M., & Jantz, R. L. 2000. Secular change in cranial facial morphology. *American Journal of Human Biology*, 12(3), 327-338. [https://doi.org/10.1002/\(SICI\)1520-6300\(200005/06\)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1520-6300(200005/06)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1)
- Jantz, R. L. 2019. Title discoveries from the Forensic Anthropology Data Base: Modern American skeletal change & the case of Amelia Earhart. *The FASEB Journal*, 33:202.1-202.1. https://doi.org/10.1096/fasebj.2019.33.1_supplement.202.1
- Jantz, R. L., & Meadows Jantz, L. 2016. The remarkable change in Euro-American cranial shape and size. *Human Biology* 88(1): 56-64.
- Jantz, R. L., Meadows Jantz, L., & Devlin, J. L. 2016. Secular changes in the postcranial skeleton of American Whites. *Human Biology*, 88(1), 65-75. <https://doi.org/10.13110/humanbiology.88.1.0065>

- Kamnkar, K. R., Hefner, J. T., Monslave, T., & Florez, L. M. B. 2021. Craniometric variation in a regional sample from Antioquia, Medellín, Colombia: Implications for forensic work in the Americas. *Forensic Anthropology* 5(3):199-210. <https://doi.org/10.5744/fa.2020.2023>
- Komar, D., & Grivas, D. 2008. Manufactured populations. What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum Documented Collection. *American Journal of Biological Anthropology* 137(2): 244-233. <https://doi.org/10.1002/ajpa.20858>
- Langley, N. R., Jantz, R. L., & Ousley, S. D. 2016. The effect of novel environments on modern American skeletons. *Human Biology* 88(1): 5-13.
- Mann, R. W., Labrash, S., & Lozanoff, S. 2020. A new osteological resource at the John A. Burn's School of Medicine. *Hawai'i Journal of Health & Social Welfare* 79(6): 202-203.
- Moore-Jansen, P. H. 1989. A multivariate craniometric analysis of secular change and variation among recent North American populations. *Unpublished doctoral dissertation*. University of Tennessee, Knoxville.
- Muller, J. L., Pearlstein, K. E., & de la Cova, C. 2017. Dissection and documented skeletal collections: Embodiments of legalized inequality. In: K. C. Nystrom, ed. *The bioarchaeology of dissection and autopsy in the United States*. New Paltz, NY: Springer. pp. 185-201.
- Ousley SD. 2014. *3Skull version 1.76*, <https://www.statemachine.net/software/3Skull/>
- R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>
- Relethford, J. H. 2009. Race and global patterns of phenotypic variation. *American Journal of Biological Anthropology*, 139(1): 16-22. <https://doi.org/10.1002/ajpa.20900>
- Sadler, A. M., Sadler, B. L., Stason, E. B. 1968. The Uniform Anatomical Gift Act: A model for reform. *JAMA* 206(1):2501-2506. doi: 10.1001/jama.1968.03150110049007
- Sharma, G. 2017. Pros and cons of different sampling techniques. *International Journal of Applied Research* 3(7):749-752.
- Sharman, J., & Albanese, J. 2018. Bioarchaeology and identified skeletal collections: Problems and potential solutions. In: C. Y. Henderson & F. A. Cardoso, eds. *Identified skeletal collections: The testing ground of anthropology?* ...
- Singh, A. S., & Masuku, M. B. 2014. Sampling techniques and determination of sample size in applied statistics research: An overview. *International Journal of Economics, Commerce and Management* 2(11):1-22.

- Smith, J. R. 2019. Living with observational data in biological anthropology. *American Journal of Biological Anthropology* 169: 591-598.
- Smith, J., & Noble, H. 2014. Bias in research. *Evidence Based Nursing* 17(4): 100-101.
- Sparks, C. S., & Jantz, R. L. 2002. A reassessment of human cranial plasticity. *PNAS*, 99(23): 14636-14639. <https://doi.org/10.1073/pnas.222389599>
- Spradley, M. K. 2008. Biological anthropological aspects of the African diaspora: Geographic origins, secular trends, and plastic versus genetic influences utilizing craniometric data. *Unpublished doctoral dissertation*. University of Tennessee, Knoxville.
- Spradley, M. K., Stull, K. E., & Hefner, J. T. 2016. Craniofacial secular change in Mexican migrants. *Human Biology* 88(1): 15-29.
- STAFS Research. 2024. Southeast Texas Applied Forensic Science Facility, Sam Houston State University. Retrieved October 13, 2024, from <https://ifrti.org/STAFS/research.html>
- Utermohle, C. J., & Zegura, S. L. (1982). Intra- and interobserver error in craniometry: A cautionary tale. *American Journal of Biological Anthropology*, 57(3): 303-310. <https://doi.org/10.1002/ajpa.1330570307>
- Utermohle, C. J., Zegura, S. L., & Heathcote, G. M. (1983). Multiple observers, humidity, and choice of precision statistics: Factors influencing craniometric data quality. *American Journal of Biological Anthropology*, 61(1): 85-95. <http://dx.doi.org/10.1002/ajpa.1330610109>
- van Buuren S, Groothuis-Oudshoorn K. Multiple imputation by chained equations (MICE) v.2.5; 2011.
- Winburn, A. P., Jennings, A. L., Steadman, D. W., & DiGangi, E. A. 2022. Ancestral diversity in skeletal collections: Perspectives on African American body donation. *Forensic Anthropology* 5(2):141-152. <https://doi.org/10.5744/fa.2020.1023>

MANUSCRIPT 2. REPRESENTATION OF DOCUMENTED SKELETAL COLLECTIONS TO FORENSIC CASEWORK

INTRODUCTION

Documented skeletal collections are invaluable data sources for forensic anthropological research (Campanacho et al. 2021); however, numerous studies have identified that modern collections in the United States are not representative of the living population (e.g., Komar and Grivas 2008; Winburn et al. 2021). This sample bias is evident in the demographic compositions of documented skeletal collections, which are often *overrepresented* by older, European American males, when compared to state or national census populations (George et al. 2021; Gocha et al. 2021; Komar and Grivas 2008; Winburn et al. 2021).

The forensic case population is made up of decedents whose circumstances of death bring them under the jurisdiction of the medicolegal death investigation system (Christensen and Passalacqua 2018). Of those cases, forensic anthropologists consult on cases where remains are in advanced stages of decomposition, including skeletonization, or require specialized trauma analysis. As these analyses often include only a subset of the general population, we need to understand if documented skeletal collections adequately represent forensic anthropology cases and if they are appropriate samples for research endeavors. One approach to answer this question is a comparison of the documented skeletal collections used in method development to the demographic composition of forensic cases.

The Forensic Anthropology Data Bank (FDB) is a database of over 4,000 individuals with associated skeletal and demographic information, including skeletal measurements, observations of various nonmetric traits, population/social race, biological sex, and year of birth (Jantz 2019). The FDB was developed to provide a more appropriate source of skeletal data for modern forensic

casework, since historic collections are known to be biased by limited population variability, influences of secular change on skeletal morphology, and an overrepresentation of older individuals from lower socioeconomic status groups (Jantz and Moore-Jansen 1987). The FDB avoids these sources of bias through the solicitation of skeletal data from identified forensic cases, ensuring these data are *representative of the population that forensic anthropologists routinely work to identify* (emphasis added; Jantz and Moore-Jansen 1987). The FDB does contain samples from documented skeletal collections, including the Robert J. Terry Collection, Maxwell Museum's Donation Skeletal Collection (MMDSC), and the University of Tennessee, Knoxville, Donated Skeletal Collection (UTK DSC), but the FDB only includes individuals from those collections born in the 20th century or later (Jantz and Ousley 2005).

One of the most significant outcomes of the FDB is the computer software FORDISC, a program facilitating the metric estimation of population affinity, sex, and stature using discriminant function analysis and linear regressions (Jantz and Ousley 2005, 2012; Ousley and Jantz 2012). FORDISC incorporates data from the FDB as reference samples. In its current version, Fordisc 3.1.322 (Fordisc or FD3) houses over 2,500 individuals from the FDB across 13 population/sex-specific reference groups (Jantz and Ousley 2005). Its automated user interface and ease of use has led to adoption worldwide, with a recent survey identifying FORDISC as the leading software program for population affinity estimation in the United States and Europe (Davidson and Morgan 2022).

The reference samples housed in FD3 provide an opportunity to assess if documented skeletal collections are representative of the forensic case population. Using the custom database feature, we can create unique reference samples derived from the FD3 to evaluate how classification accuracies and group relationships change as documented skeletal collections are

added to (or removed from) the reference samples in FD3. This study investigates the representativeness of modern documented skeletal collections to the forensic anthropology case population using samples from eight collections within the United States and the Fordisc 3.1.322 reference sample data.

MATERIALS AND METHODS

Samples

Data used in this study include craniometric measurements from 2,612 individuals, including a subset of individuals from Fordisc 3.1.322 and author (RRD) collected samples from eight United States-based documented skeletal collections (DSCs). Craniometric data have been used as a proxy for genetic markers due to the high heritability of cranial morphology and a demonstrable correlation to genetic variation. Numerous studies have shown how population variability (in cranial morphology) often closely follows patterns observed in genetic data, reflecting shared ancestry and gene flow (along with other aspects of genetics like isolation by distance). While environmental and cultural factors can influence cranial form, the robust genetic component of cranial morphology—validated through comparisons with DNA (Herrera et al. 2014; Relethford 1994; Roseman 2004; Roseman and Weaver 2004)—provide a very reliable tool for investigating population history and biological distance in the absence of genetic data (e.g., Jantz and Jantz, 2000; Relethford, 2009; Sparks and Jantz, 2002). These factors, as well as anthropometric patterns not attributable to general human variation (Gordon and Bradtmiller 1992; Utermohle and Zegura 1982; Utermohle et al. 1983), provide cranial morphology with a unique ability to capture population-specific signatures, offering valuable insights into both genetic relationships and the evolutionary forces shaping human diversity.

Twenty-two cranial measurements (Table 2.1) from 1,564 individuals in Fordisc across all

thirteen population/sex reference groups (Table 2.2) are used in the following analyses. During each individual analysis, Fordisc filters reference samples to include individuals with all recorded measurements (i.e., no missing data). These 22 measurements were selected to maximize reference sample sizes and to comply with the recommendations in Fordisc regarding the maximum number of measurements (i.e., the number of variables should not exceed the smallest sample size, minus one; Jantz and Ousley 2005). In the end, three datasets were created from the Fordisc sample: 1) *Complete* - all FD3 data ($n = 1,564$); 2) *No MMDSC* – the FD3 dataset with the Maxwell Museum’s Documented Skeletal Collection (MMDSC) removed ($n = 1,535$); and, 3) *No UTK DSC* - the FD3 dataset with the UTK DSC removed ($n = 1,316$; see Table 2.2).

Table 2.1. Cranial measurements (and abbreviations) used throughout the analyses.

<i>Abbreviation</i>	<i>Measurement</i>	<i>Abbreviation</i>	<i>Measurement</i>
<i>GOL</i>	Cranial length	<i>MAL</i>	Max. Alveolar length
<i>BNL</i>	Cranial base length	<i>MDH</i>	Mastoid Height
<i>BBH</i>	Basion-bregma height	<i>OBH</i>	Orbital height
<i>XCB</i>	Cranial breadth	<i>OBB</i>	Orbital breadth
<i>WFB</i>	Minimum frontal breadth	<i>DKB</i>	Interorbital breadth
<i>ZYB</i>	Bizygomatic breadth	<i>EKB</i>	Biorbital breadth
<i>AUB</i>	Biauricular breadth	<i>FRC</i>	Frontal chord
<i>BPL</i>	Basion-prosthion length	<i>PAC</i>	Parietal chord
<i>NLH</i>	Nasal height	<i>OCC</i>	Occipital chord
<i>NLB</i>	Nasal breadth	<i>FOL</i>	Foramen magnum length
<i>MAB</i>	Max. Alveolar breadth	<i>UFHT</i>	Upper facial height

Table 2.2. Breakdown of the three FD3 subsamples used in this study, by population and sex.

	<i>Complete Sample (n)</i>		<i>No MMDSC (n)</i>			<i>No UTK DSC (n)</i>		
	M	F	M	F	% of Total	M	F	% of Total
African American (B)	138	93	138	93	100.0	110	88	85.7
Chinese (CH)	73	0	73	0	100.0	73	0	100.0
European American (W)	337	218	318	214	95.9	193	155	62.7
Guatemalan (GT)	67	0	67	0	100.0	67	0	100.0
Hispanic (H)	179	36	175	36	98.1	171	36	96.3
Japanese (J)	183	113	183	113	100.0	183	113	100.0
Native American (A)	51	28	49	28	97.5	51	28	100.0
Vietnamese (V)	48	0	48	0	100.0	48	0	100.0
Total	1076	488	1051	484	98.1	896	420	84.1

Note: Fordisc population abbreviations included in parentheses.

The DSC subsample includes 1,048 individuals from eight modern documented skeletal collections located within the United States (Table 2.3). Two of these collections—MMDSC and UTK DSC—were chosen because they are included in current FDB and Fordisc reference samples. Six other collections were included to provide national representation of individuals housed in modern, documented skeletal collections, including collections associated with outdoor decomposition facilities, medical schools, and/or medical examiner offices. These include: 1) the *John A. Williams Documented Human Skeletal Collection* (JAWDHSC); 2) the *Mann-Labrash Osteological Collection* (MLOC); 3) the *Michigan State University Forensic Anthropology Laboratory Donated Skeletal Collection* (MSUFAL DSC); 4) the *Southeast Texas Applied Forensic Science Facility Donated Skeletal Collection* (STAFS DSC); 5) the *Texas State Donated Skeletal Collection* (TXSTDSC); and, 6) the *Western Michigan University's Homer Stryker, M.D., School of Medicine Skeletal Teaching and Research Series* (WMed STARS; Table 2.4). A simple random sampling strategy during data collection ensured adequate collection representation; however, this selection process resulted in unbalanced population and sex samples. Only those individuals with a population and sex classification matching those in the Fordisc reference samples were included in this analysis. For MMDSC and UTK DSC, those individuals already included in the Fordisc reference sample were excluded from the DSC sample.

Craniometric data for the collection sample were collected by the first author (RRD) using a Microscribe® G digitizer and the 3Skull software program (Ousley 2014).

Table 2.3. Breakdown of collection sample (by population and sex). A breakdown of the subset of collection data containing complete observations for all 22 measurements is also provided.

	Total (n)		Complete Observations (n)	
	M	F	M	F
<i>John A. Williams Documented Human Skeletal Collection (n = 118, all measurements present = 37)</i>				
African American (B)	1	0	1	0
European American (W)	58	59	18	18
<i>Mann-Labrash Osteological Collection (n = 77, all measurements present = 32)</i>				
African American (B)	3	0	1	0
Chinese (CH)	4	0	2	0
European American (W)	31	24	11	11
Japanese (J)	7	7	4	2
Vietnamese (V)	1	0	1	0
<i>Maxwell Museum's Documented Skeletal Collection (n = 209, all measurements present = 53)</i>				
African American (B)	3	1	0	0
European American (W)	106	89	29	22
Hispanic (H)	9	1	2	0
<i>MSUFAL Donated Skeletal Collection (n = 36, all measurements present = 15)</i>				
African American (B)	5	4	2	2
European American (W)	20	7	10	1
<i>STAFS Donated Skeletal Collection (n = 149, all measurements present = 20)</i>				
African American (B)	7	1	0	0
European American (W)	81	51	14	2
Hispanic (H)	6	2	3	0
Native American (A)	1	0	1	0
<i>Texas State Donated Skeletal Collection (n = 239, all measurements present = 69)</i>				
African American (B)	0	3	0	2
European American (W)	129	94	39	26
Hispanic (H)	7	5	1	1
Native American (A)	0	1	0	0
<i>UTK Donated Skeletal Collection (n = 182, all measurements present = 30)</i>				
European American (W)	93	88	12	17
Hispanic (H)	0	1	0	1
<i>WMed Skeletal Teaching and Research Series (n = 38, all measurements present = 10)</i>				
African American (B)	0	1	0	0
European American (W)	26	9	6	3
Native American (A)	1	1	1	0
Total	599	449	158	108

Note: Fordisc population abbreviations included in parentheses.

Table 2.4. Location, year of establishment, collection total, and primary donor sources for each documented skeletal collection.

<i>Collection</i>	<i>Location</i>	<i>Established</i>	<i>Collection Total (n)*</i>	<i>Primary Donor Sources</i>
JAWDHSC	Cullowhee, NC	2003	129	Outdoor decomposition facility (George et al. 2022)
MLOC	Honolulu, HI	2019	~350	Medical school (Mann et al. 2020)
MMDSC	Albuquerque, NM	1978	314	Office of the Medical Investigator (Komar and Grivas 2008)
MSUFAL DSC	East Lansing, MI	1996	42	Medical examiners offices
STAFS DSC	Huntsville, TX	2009	~450	Outdoor decomposition facility (STAFS Research 2024)
TXSTDSC	San Marcos, TX	2008	~700	Outdoor decomposition facility (Gocha et al. 2022)
UTK DSC	Knoxville, TN	1981	~1800	Outdoor decomposition facility (Winburn et al. 2020)
WMed STARS	Kalamazoo, MI	2014	53	Medical school and Office of the Medical Examiner

*Exact numbers accurate as of data collection trip.

Of the DSC sample, 266 individuals had complete observations for all 22 cranial measurements (see Tables 2.1 and 2.3). To bolster sample size, missing values were imputed for the remaining 782 individuals, resulting in: 1) a subset of DSC data with all observations ($n = 266$); and, 2) a larger dataset with missing values imputed, by variable ($n = 1,048$; see Table 2.3). All data were imputed using the ‘*mice*’ package in R and the ‘predictive mean matching’ approach (R Core Team 2024).

Craniometric Analysis

All data were pooled and used in various combinations to generate 57 custom datasets (Table 2.5). The purpose of these unique datasets was to explore the impact larger samples of donated skeletal collections have on the results generated in Fordisc.

Table 2.5. Sample counts for each custom database.

Collections Used in Database	FD3 Sample Used		
	Complete FD3 Dataset (n)	No MMDSC (n)	No UTK DSC (n)
No Collections	1564	1535	1316
<i>Collection Data Subsets with Complete Observations</i>			
All Collections	1830	1801	1582
JAWDHSC	1601	1572	1353
MLOC	1596	1567	1348
MMDSC	1617	1588	1369
MSUFAL DSC	1579	1550	1331
STAFS DSC	1584	1555	1336
TXSTDSC	1633	1604	1385
UTK DSC	1594	1565	1346
WMed STARS	1574	1545	1326
<i>Collection Datasets with Missing Values Imputed</i>			
All Collections	2612	2583	2364
JAWDHSC	1682	1653	1434
MLOC	1641	1612	1393
MMDSC	1773	1744	1525
MSUFAL DSC	1600	1571	1352
STAFS DSC	1713	1684	1465
TXSTDSC	1803	1774	1555
UTK DSC	1746	1717	1498
WMed STARS	1602	1573	1354

For each analysis, data was imported into Fordisc using the “Custom Database” feature. Prior to processing, “Classify Case” was switched off in the “Options” tab, and the “Group VCVMs” (or group variance-covariance matrices) and “Individual Scores” options were engaged. All 22 cranial measurements were selected for each analysis and Population/Sex (i.e., “PopSex”) was chosen as the grouping variable. Following each iteration, detected outliers were removed using the remove outlier function in Fordisc; data was re-processed until no outliers remained. Overall correct classifications, each group’s variance-covariance matrix (VCVM), and individual classifications were saved for each analysis.

Fordisc classifies unknown individuals using linear discriminant function analysis. Following measurement and reference group selection, discriminant functions are created by assigning numerical weights to each predictor variable (i.e., cranial measurements) which

maximize differences between groups and minimize the differences within groups (Ousley and Jantz 2012). The number of discriminant functions created depends on the number of reference groups and/or predictor variables; one fewer discriminant function than reference groups are produced unless there are fewer predictor variables than reference groups (Kachigan 1982). An unknown individual is compared to reference group centroids via a Mahalanobis distance and is classified into the most similar (closest) group. Summary statistics, including posterior probabilities and various typicality measures, are provided following classification. When 'Classify Case' is turned off, only overall group classification accuracies are calculated, using a cross-validated (leave-one-out; LOOCV) procedure. Fordisc classifies all individuals, regardless of true group membership; therefore, the user must assess whether the appropriate reference samples for a particular case are included (c.f., Birkby 1966).

To evaluate the representativeness of the collections, the dataset comprising the complete Fordisc sample, including all collections with imputed missing values, was reanalyzed using three grouping variables: source and population (pooled sexes with shape transformation applied); source and sex; and, source, population, and sex. In this context, "source" refers to the dataset's origin (FD3, UTK DSC, MMDSC, etc.). For each analysis, Fordisc generated plots using canonical variates analysis (CVA) and the calculated Mahalanobis distances. These plots were reviewed to assess group separation and examine the relationships among groups.

RESULTS

Initial Data Analysis

Some data cleaning was necessary to combine the Fordisc and DSC datasets. All of these changes were related to data formatting and structure. For example, demographic information from the DSC sample was formatted to match the Fordisc dataset (i.e., population abbreviations were

changed to match Fordisc). The Fordisc and DSC samples were pooled and provided unique identifiers so individual classifications could be examined using the Extended Results tab in Fordisc.

The demographic composition differs across data sources, as shown in Table 6. The individuals in the documented skeletal collections are primarily European Americans; Fordisc has, by far, the greatest population variability (i.e., number and proportion of population groups represented). However, population variability of each collection sample slightly increases after missing values are imputed.

Table 2.6. Demographic breakdown of FD3 and Collection samples.

Group	Non-Imputed Samples																	
	FD3		JAWDHS C		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC C		UTK DSC		WMed STARS	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
AF	28	1.8																
AM	51	3.3									1	5.0					1	10.0
BF	93	5.9							2	13.3			2	2.9				
BM	138	8.8	1	2.7	1	3.1			2	13.3								
CHM	73	4.7			2	6.3												
GTM	67	4.3																
HF	36	2.3					2	3.8					1	1.4	1	3.3		
HM	179	11.4									3	15.0	1	1.4				
JF	113	7.2			2	6.3												
JM	183	11.7			4	12.5												
VM	48	3.1			1	3.1												
			1		1								2					
WF	218	13.9	8	48.6	1	34.4	22	41.5	1	6.7	2	10.0	6	37.7	17	56.7	3	30.0
			1		1								3					
WM	337	21.5	8	48.6	1	34.4	29	54.7	10	66.7	14	70.0	9	56.5	12	40.0	6	60.0
	156	100.	3		3	100.							6	100.				
Total	4	0	7	100.0	2	0	53	100.0	15	100	20	100.0	9	0	30	100.0	10	100.0

Group	Imputed Samples															
	JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
AF											1	0.4			1	2.6
AM									1	0.7					1	2.6
BF					1	0.5	4	11.1	1	0.7	3	1.3			1	2.6
BM	1	0.8	3	3.9	3	1.4	5	13.9	7	4.7						
CHM			4	5.2												
GTM																
HF					1	0.5			2	1.3	5	2.1	1	0.5		
HM					9	4.3			6	4.0	7	2.9				
JF			7	9.1												
JM			7	9.1												
VM			1	1.3												
			2		8								8			
WF	59	50.0	4	31.2	9	42.6	7	19.4	51	34.2	94	39.3	8	48.4	9	23.7

Table 2.6. (cont'd)

WM	58	49.2	31	40.3	106	50.7	20	55.6	81	54.4	129	54.0	93	51.1	26	68.4
Total	118	100.0	77	100.0	209	100.0	36	100.0	149	100.0	239	100.0	182	100.0	38	100.0

Cranial measurements with the highest counts of missing observations were those associated with prosthion and ectomalare (Figure 2.1), likely due to antemortem tooth loss and alveolar resorption. Following imputation, variable means were calculated (Table 2.7).

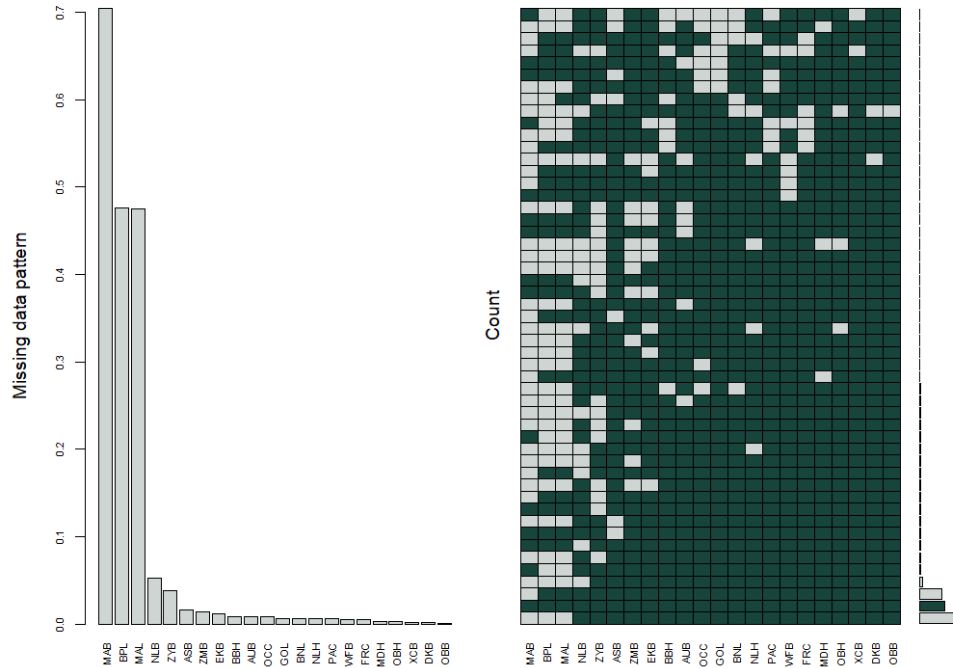


Figure 2.1. Visualization of missing data patterns among cranial measurements across all documented skeletal collection samples. Missing values (grey) and observed values (green) are provided for each variable.

Table 2.7. Mean measurements for each data source: Fordisc and each documented skeletal collection (both non-imputed and imputed samples).

<i>Data</i>	<i>Collection</i>	<i>n</i>	GOL			BNL			BBH			XCB			WFB		
			<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
Non-Imputed	FD3	1564	180	8.8	57	101	5.4	33	136	6.4	41	138	6.0	43	94	5.0	39
	JAWDHSC	37	185	8.5	34	104	5.4	27	139	6.2	23	138	6.0	26	96	5.5	25
	MLOC	32	183	8.1	32	104	4.4	19	140	4.9	21	142	7.5	30	96	5.1	22
	MMDSC	53	183	7.5	30	103	6.1	32	136	6.6	28	138	6.0	28	95	5.1	24
	MSUFAL DSC	15	186	8.2	26	105	4.4	14	140	7.7	28	135	4.7	16	94	5.7	19
	STAFS DSC	20	187	8.2	32	107	5.0	19	142	5.4	20	140	6.0	25	97	3.7	17
	TXSTDSC	69	185	8.2	41	104	5.9	29	139	6.4	35	139	5.8	27	94	5.0	23
	UTK DSC	30	185	9.8	37	102	5.0	17	137	6.9	27	141	5.5	23	94	4.8	20
Imputed	WMed STARS	10	185	6.3	23	104	3.4	10	139	5.0	15	138	3.7	11	95	4.1	12
	JAWDHSC	118	184	7.8	34	103	5.0	27	138	6.2	32	138	5.5	27	95	4.9	25
	MLOC	77	183	8.4	40	103	5.2	26	138	5.4	29	141	6.4	33	95	5.3	25
	MMDSC	209	183	8.1	51	102	6.0	32	136	6.3	29	138	5.5	30	94	5.0	27
	MSUFAL DSC	36	183	9.5	33	103	6.0	23	137	7.7	32	136	5.5	22	94	6.0	25
	STAFS DSC	149	183	8.7	53	104	5.8	34	138	6.4	34	138	5.9	35	95	4.7	22
	TXSTDSC	239	184	8.2	45	103	5.5	29	138	6.2	35	139	5.6	32	95	4.8	24
	UTK DSC	182	183	8.4	46	102	5.3	24	136	6.3	30	139	5.8	33	94	5.1	33
	WMed STARS	38	185	7.4	27	104	4.9	20	139	6.2	23	138	6.1	25	96	5.2	24
<i>Data</i>	<i>Collection</i>	<i>n</i>	ZYB			AUB			BPL			NLH			NLB		
			<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
Non-Imputed	FD3	1564	129	6.8	47	122	5.9	37	97	6	42	51	3.4	22	25	2.2	16
	JAWDHSC	37	125	5.7	27	120	5.6	25	97	5.9	26	53	3.7	16	23	2.3	9
	MLOC	32	129	7.1	31	125	7.4	32	97	6.6	33	54	3.5	14	24	2.7	12
	MMDSC	53	125	6.7	25	120	5.6	23	95	6.4	28	53	3.6	16	24	2.4	12
	MSUFAL DSC	15	123	5.4	19	118	5.1	14	100	5.8	15	52	2.2	8	23	1.5	5
	STAFS DSC	20	128	5.4	23	124	5.3	22	100	5.3	21	55	3.6	14	24	1.8	7
	TXSTDSC	69	125	6.4	28	121	5.4	22	97	5.8	27	53	3.0	15	23	2.0	12
	UTK DSC	30	125	5.0	22	122	5.3	22	95	6.6	26	53	3.4	16	23	1.7	6

Table 2.7. (cont'd)

Imputed	WMed STARS	10	127	6.2	22	122	4.7	16	99	6.1	19	52	2.6	9	24	2.5	8
	JAWDHSC	118	124	5.8	29	120	5.1	26	96	5.5	30	53	3.5	16	23	2.1	11
	MLOC	77	128	6.5	33	124	6.2	32	97	6.4	33	53	3.3	15	25	2.4	12
	MMDSC	209	125	6.8	42	120	5.7	38	95	6.8	36	54	3.8	19	24	2.2	13
	MSUFAL DSC	36	124	5.5	23	119	4.9	18	97	6.7	31	53	3.0	14	24	2.2	10
	STAFS DSC	149	126	6.1	29	122	5.1	27	97	6.5	39	54	3.4	17	24	2.0	13
	TXSTDSC	239	125	6.2	30	121	5.5	29	96	6.0	33	53	3.5	20	24	1.8	12
	UTK DSC	182	126	5.9	25	122	5.5	29	94	6.4	31	53	3.5	18	24	2.2	11
	WMed STARS	38	125	6.5	27	121	6.0	31	97	5.9	26	53	2.5	11	24	2.0	8

Data	Collection	n	MAB				MAL			MDH			OBH			OBB		
			mean	sd	range	mean	sd	range	mean	sd	range	mean	sd	range	mean	sd	range	
Non-Imputed	FD3	1564	63	4.7	30	54	3.9	24	30	4.3	31	34	2.2	22	40	2.3	14	
	JAWDHSC	37	60	4.8	22	55	3.8	14	28	3.2	15	34	2.0	8	41	1.8	7	
	MLOC	32	62	6.2	26	55	4.2	19	28	4.0	16	35	2.4	10	41	2.2	8	
	MMDSC	53	60	4.4	17	54	3.7	14	29	3.3	14	34	1.9	8	40	1.8	9	
	MSUFAL DSC	15	61	3.1	12	55	3.4	12	29	3.7	12	34	1.0	3	41	1.7	6	
	STAFS DSC	20	63	4.6	14	56	3.9	14	30	3.5	17	34	1.6	7	41	1.8	6	
	TXSTDSC	69	61	4.0	20	54	3.3	14	28	3.4	16	34	1.8	8	40	2.0	11	
	UTK DSC	30	60	4.0	18	53	4.1	18	27	3.9	15	34	1.7	6	41	2.0	9	
	WMed STARS	10	60	3.6	12	55	2.2	7	30	4.0	13	33	1.7	5	41	1.6	5	
Imputed	JAWDHSC	118	60	4.2	22	54	3.4	16	28	3.2	18	34	1.9	8	41	2.0	10	
	MLOC	77	62	5.4	26	54	3.7	20	28	3.6	18	35	2.3	12	41	2.0	8	
	MMDSC	209	61	4.7	24	54	4.2	28	29	3.3	17	34	1.9	11	40	2.1	10	
	MSUFAL DSC	36	61	4.4	20	55	3.9	20	28	3.8	14	34	1.8	8	41	2.4	13	
	STAFS DSC	149	62	5.1	23	54	3.8	20	29	3.4	17	34	2.0	10	41	1.8	11	
	TXSTDSC	239	61	4.4	23	54	3.7	22	28	3.7	17	34	2.0	10	41	2.0	11	
	UTK DSC	182	60	4.3	23	53	4.1	25	28	3.6	17	34	2.0	10	41	2.0	10	
	WMed STARS	38	60	4.0	17	54	3.1	12	29	3.4	14	34	2.0	10	41	2.0	9	

Table 2.7. (cont'd)

<i>Data</i>	<i>Collection</i>	<i>n</i>	DKB			EKB			FRC			PAC			OCC		
			<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>
Non-Imputed	FD3	1564	21	2.6	19	97	4.3	31	111	5.5	36	114	7.2	44	98	5.9	37
	JAWDHSC	37	20	2.2	9	97	4.4	17	113	5.0	23	116	5.9	27	99	5.7	25
	MLOC	32	21	2.2	10	98	5.3	21	114	5.2	20	115	6.8	29	99	5.5	23
	MMDSC	53	21	2.0	8	96	4.1	20	112	5.0	26	115	5.9	27	97	5.7	34
	MSUFAL DSC	15	20	2.2	8	95	3.6	13	113	5.3	16	115	6.7	19	101	7.6	27
	STAFS DSC	20	21	2.0	7	97	3.9	14	116	5.8	22	117	8.1	30	101	5.8	24
	TXSTDSC	69	20	2.2	10	96	4.1	18	114	5.8	31	116	5.8	26	99	4.7	20
	UTK DSC	30	20	1.9	7	96	4.3	19	113	6.8	28	116	7.8	29	99	8.2	35
	WMed STARS	10	21	2.1	6	97	3.1	9	112	3.6	12	120	5.3	15	97	5.3	19
Imputed	JAWDHSC	118	20	2.2	11	96	4.2	19	112	5.0	28	116	6.7	34	99	5.8	32
	MLOC	77	21	2.1	10	98	4.6	21	113	4.9	28	115	7.0	36	99	5.7	25
	MMDSC	209	21	2.2	12	96	4.4	22	111	5.7	34	115	6.3	35	97	5.1	36
	MSUFAL DSC	36	20	2.6	13	96	4.2	19	112	5.7	24	115	6.7	28	99	6.9	35
	STAFS DSC	149	21	2.0	12	97	3.6	24	113	5.1	24	115	7.5	36	99	6.3	36
	TXSTDSC	239	20	2.1	13	96	3.9	19	113	5.4	31	116	6.8	35	98	5.2	32
	UTK DSC	182	20	2.1	12	96	4.2	25	112	5.5	28	115	6.5	32	98	5.7	37
	WMed STARS	38	21	2.1	8	97	3.9	18	113	5.3	20	117	5.7	23	99	5.4	19
<i>Data</i>	<i>Collection</i>	<i>n</i>	FOL			UFHT											
			<i>mean</i>	<i>sd</i>	<i>range</i>	<i>mean</i>	<i>sd</i>	<i>range</i>									
Non-Imputed	FD3	1564	36	2.6	19	71	4.7	33									
	JAWDHSC	37	37	2.6	11	71	5.7	27									
	MLOC	32	37	2.0	8	72	4.3	18									
	MMDSC	53	38	2.6	10	71	4.7	21									
	MSUFAL DSC	15	36	2.5	8	71	3.3	10									
	STAFS DSC	20	37	2.4	10	72	4.5	18									
	TXSTDSC	69	38	2.5	10	72	4.4	21									
	UTK DSC	30	37	2.6	12	70	4.8	23									
	WMed STARS	10	37	2.0	7	71	4.3	15									

Table 2.7. (cont'd)

Imputed	JAWDHSC	118	37	2.7	18	71	5.1	27
	MLOC	77	37	2.4	11	71	4.0	20
	MMDSC	209	37	2.6	14	71	4.9	27
	MSUFAL DSC	36	37	2.7	12	70	4.3	21
	STAFS DSC	149	37	2.9	18	71	4.8	29
	TXSTDSC	239	37	2.5	14	71	4.8	28
	UTK DSC	182	37	2.7	13	70	4.9	26
	WMed STARS	38	37	2.6	10	70	3.8	17

Craniometric Analysis

All discriminant functions produced robust group separations (Wilks' Lambda, $p < 0.0001$). Across all 57 analyses, VCVMs showed no significant differences between groups (VCVM Homogeneity test, $p = 1.00$), and all data met the assumptions required by linear discriminant function analysis.

Overall correct classifications are provided in Table 2.8. Classification rates are highest using the imputed data. The lowest rates are associated with datasets containing the Fordisc sample without the UTK DSC individuals. Correct classification rates for each population and sex are provided in Table 2.9. European American males and females consistently attain the highest correct classification rates across all models.

Table 2.8. Overall LOOCV correct classification rates of population and sex groups for models built using each dataset. Sample sizes account for any outliers removed during analysis.

Collections Used in Database	Full FD3 Dataset		FD3 without MMDSC		FD3 without UTK DSC	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
No Collections	1536	61.3	1506	61.1	1292	59.3
<i>Collection Data Subsets with Complete Observations</i>						
All Collections	1799	63.6	1770	63.6	1555	62.1
JAWDHSC	1574	62.1	1545	61.4	1330	59.8
MLOC	1566	61.4	1537	61.2	1324	59.5
MMDSC	1590	61.9	1561	61.5	1344	60.2
MSUFAL DSC	1551	61.5	1521	61.1	1306	59.8
STAFS DSC	1554	61.5	1525	61.4	1311	59.3
TXSTDSC	1604	62.5	1574	62.3	1359	60.6
UTK DSC	1564	62.0	1535	61.6	1320	59.8
WMed STARS	1546	61.4	1516	61.1	1301	59.2
Average		62.0		61.7		60.0
<i>Collection Datasets with Missing Values Imputed</i>						
All Collections	2567	66.3	2567	66.3	2320	65.9
JAWDHSC	1653	62.9	1653	62.9	1409	61.1
MLOC	1612	60.9	1612	60.9	1367	59.3
MMDSC	1741	62.8	1741	62.8	1498	61.1
MSUFAL DSC	1570	61.7	1570	61.7	1328	59.9
STAFS DSC	1682	62.5	1682	62.5	1436	60.4
TXSTDSC	1774	63.6	1774	63.6	1525	61.9
UTK DSC	1713	63.3	1713	63.3	1468	61.1
WMed STARS	1572	61.6	1572	61.6	1329	59.7
Average		62.8		62.8		61.2

Table 2.9. Sample sizes and correct classification rates for population/sex groups in models from each of the 57 datasets.

Full FD3 Sample and Non-Imputed Collection Samples																				
Full FD3			All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
Group	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR
AF	27	55.6	27	63	27	59.3	27	55.6	27	59.3	27	59.3	27	59.3	27	59.3	27	55.6	27	55.6
AM	47	63.8	47	61.2	47	63.8	47	63.8	47	63.8	47	63.8	48	62.5	47	63.8	47	63.8	48	60.4
BF	88	62.5	93	65.6	89	62.9	88	64.8	89	61.8	90	62.2	88	63.6	90	65.6	88	64.8	88	62.5
BM	131	62.6	134	64.2	132	62.1	131	61.8	131	62.6	133	62.4	131	63.4	130	61.5	131	63.4	131	62.6
CHM	73	47.9	75	49.3	73	47.9	75	46.7	73	47.9	73	47.9	73	47.9	73	47.9	73	47.9	73	47.9
GTM	66	69.7	66	68.2	66	69.7	66	68.2	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7
HF	36	58.3	38	57.9	36	58.3	36	58.3	36	58.3	36	58.3	36	58.3	37	59.5	37	56.8	36	58.3
HM	175	30.9	180	30	175	31.4	174	30.5	177	31.1	175	30.3	177	30.5	176	31.3	174	31	175	30.9
JF	113	63.7	115	62.6	113	64.6	115	62.6	113	63.7	113	63.7	113	61.9	113	62.8	113	63.7	113	63.7
JM	183	44.8	187	44.4	183	47	187	45.5	183	45.4	183	45.4	183	44.8	183	45.9	183	45.4	183	45.9
VM	47	63.8	48	62.5	47	63.8	48	62.5	47	63.8	47	63.8	47	63.8	47	63.8	47	63.8	47	63.8
WF	217	75.1	317	76.7	235	74.5	228	75.9	239	75.7	218	75.7	219	75.3	243	77.8	234	74.8	220	75.5
WM	333	77.2	470	77.7	351	77.8	344	76.7	362	77.1	343	77	346	77.2	372	76.9	344	77.9	339	76.7
Full FD3 Sample and Imputed Collection Samples																				
			All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
Group	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR	n	CCR
AF			29	62.1	27	59.3	27	59.3	27	59.3	27	59.3	28	64.3	28	60.7	27	55.6	28	53.6
AM			48	62.5	47	63.8	47	66	47	63.8	47	63.8	48	62.5	47	61.7	47	66	48	60.4
BF			99	63.6	89	62.9	89	64	90	63.3	92	65.2	90	61.1	92	64.1	88	64.8	89	64
BM			145	62.8	131	62.6	133	62.4	133	60.9	136	61.8	136	64.7	130	63.8	130	61.5	131	63.4
CHM			77	46.8	73	47.9	77	46.8	73	49.3	73	47.9	73	49.3	73	49.3	73	46.6	73	46.6
GTM			66	68.2	66	66.7	66	68.2	66	68.2	66	69.7	66	71.2	66	69.7	66	68.2	66	69.7
HF			45	64.4	36	58.3	36	58.3	37	62.2	36	58.3	38	60.5	41	56.1	37	59.5	36	58.3

Table 2.9. (cont'd)

HM	195	26.7	174	31.6	174	29.3	182	30.2	174	30.5	180	31.1	182	30.2	174	29.9	174	31
JF	120	60.8	113	62.8	120	59.2	113	61.9	113	63.7	113	64.6	113	62.8	113	63.7	113	65.5
JM	190	42.1	183	45.9	190	42.1	183	45.9	183	44.8	183	45.9	183	45.4	183	45.9	183	45.4
VM	48	62.5	47	61.7	48	64.6	47	63.8	47	61.7	47	61.7	47	63.8	47	63.8	47	63.8
WF	638	78.1	276	76.8	241	75.1	306	76.8	224	74.1	268	75	310	77.7	305	77.4	226	74.8
WM	867	75.9	391	77.7	364	76.6	437	76	352	78.1	412	75.7	462	76.8	423	77.3	358	76.5

FD3 without MMDSC and Non-Imputed Collection Samples

Group	FD3 No MMDSC		All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
AF	27	59.3	27	63	27	59.3	27	59.3	27	59.3	27	59.3	27	59.3	27	59.3	27	55.6	27	55.6
AM	45	62.2	47	57.4	45	62.2	45	62.2	45	62.2	45	62.2	46	60.9	45	62.2	45	62.2	46	60.9
BF	88	63.6	93	65.6	89	62.9	88	63.6	89	62.9	90	63.3	88	63.6	90	64.4	88	64.8	88	63.6
BM	131	61.8	134	64.2	132	61.4	131	61.8	131	62.6	133	62.4	131	62.6	130	62.3	131	62.6	131	61.8
CHM	73	47.9	75	48	73	47.9	75	46.7	73	47.9	73	46.6	73	49.3	73	49.3	73	47.9	73	46.6
GTM	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7	66	69.7	66	68.2	66	69.7
HF	36	58.3	38	57.9	36	58.3	36	58.3	36	58.3	36	58.3	36	58.3	37	59.5	37	56.8	36	58.3
HM	170	30	176	31.3	171	29.2	170	30.6	173	30.1	170	29.4	173	30.1	171	30.4	170	30.6	170	30
JF	113	63.7	115	62.6	113	63.7	115	62.6	113	61.9	113	63.7	113	62.8	113	61.9	113	61.9	113	63.7
JM	183	45.4	187	44.4	183	45.9	187	44.9	183	45.5	183	45.4	183	45.4	183	46.4	183	45.9	183	46.4
VM	47	61.7	48	62.5	47	63.8	48	62.5	47	61.7	47	61.7	47	63.8	47	61.7	47	63.8	47	61.7
WF	213	75.1	313	77	231	74	224	75	325	75.7	214	75.7	215	75.8	239	77.4	230	74.8	216	75.5
WM	314	77.1	451	77.6	332	77.7	325	77.5	343	77	324	76.5	327	77.1	353	77.1	325	78.2	320	76.6

FD3 without MMDSC and Imputed Collection Samples

Group	All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
AF	29	62.1	27	59.3	27	59.3	27	59.3	27	59.3	28	64.3	28	60.7	27	55.6	28	53.6

Table 2.9. (cont'd)

AM	48	62.5	47	63.8	47	66	47	63.8	47	63.8	48	62.5	47	61.7	47	66	48	60.4
BF	99	63.6	89	62.9	89	64	90	63.3	92	65.2	90	61.1	92	64.1	88	64.8	89	64
BM	145	62.8	131	62.6	133	62.4	133	60.9	136	61.8	136	64.7	130	63.8	130	61.5	131	63.4
CHM	77	46.8	73	47.9	77	46.8	73	49.3	73	47.9	73	49.3	73	49.3	73	46.6	73	46.6
GTM	66	68.2	66	66.7	66	68.2	66	68.2	66	69.7	66	71.2	66	69.7	66	68.2	66	69.7
HF	45	64.4	36	58.3	36	58.3	37	62.2	36	58.3	38	60.5	41	56.1	37	59.5	36	58.3
HM	195	26.7	174	31.6	174	29.3	182	30.2	174	30.5	180	31.1	182	30.2	174	29.9	174	31
JF	120	60.8	113	62.8	120	59.2	113	61.9	113	63.7	113	64.6	113	62.8	113	63.7	113	65.5
JM	190	42.1	183	45.9	190	42.1	183	45.9	183	44.8	183	45.9	183	45.4	183	45.9	183	45.4
VM	48	62.5	47	61.7	48	64.6	47	63.8	47	61.7	47	61.7	47	63.8	47	63.8	47	63.8
WF	638	78.1	276	76.8	241	75.1	306	76.8	224	74.1	268	75	310	77.7	305	77.4	226	74.8
WM	867	75.9	391	77.7	364	76.6	437	76	352	78.1	412	75.7	462	76.8	423	77.3	358	76.5

FD3 without UTK DSC and Non-Imputed Collection Samples

Group	FD3 No UTK DSC		All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
AF	27	59.3	27	63	27	63	27	59.3	27	63	27	59.3	27	63	27	63	27	59.3	27	59.3
AM	47	61.7	49	63.3	47	63.8	47	66	47	63.8	47	63.8	48	60.4	47	63.8	47	61.7	48	60.4
BF	83	63.9	88	65.9	84	64.3	84	64.3	84	63.1	85	64.7	83	63.9	86	66.3	83	65.1	83	63.9
BM	105	61	107	61.7	106	61.3	106	60.4	103	61.2	106	61.3	105	61	103	61.2	104	59.6	104	60.6
CHM	73	45.2	75	49.3	73	45.2	75	46.7	73	47.9	73	45.2	73	46.6	72	45.8	73	46.6	73	45.2
GTM	66	71.2	66	72.7	66	71.2	66	69.7	66	71.2	66	71.2	66	72.7	66	72.7	66	72.7	66	71.2
HF	36	58.3	38	60.5	36	61.1	36	58.3	36	58.3	36	58.3	36	58.3	37	59.5	37	56.8	36	58.3
HM	167	31.7	173	30.1	167	31.1	167	31.1	169	32.5	167	31.1	170	31.8	168	30.4	167	31.7	167	31.1
JF	113	62.8	115	62.6	113	62.8	115	62.6	113	61.1	113	62.8	113	61.1	113	62.8	113	62.8	113	61.9
JM	183	43.7	187	43.3	183	44.8	187	44.9	183	45.4	183	44.3	183	44.3	183	43.7	183	42.6	183	44.3
VM	47	66	48	62.5	47	66	48	64.6	47	63.8	47	63.8	47	66	47	66	47	66	47	66
WF	154	77.9	254	78	172	76.2	165	76.4	176	77.3	155	78.7	156	76.9	180	78.9	171	75.4	157	77.7

Table 2.9. (cont'd)

WM	191	77.5	328	77.1	209	77	201	77.6	220	77.3	201	78.6	204	77	230	77.4	202	80.7	197	77.2
FD3 without UTK DSC and Imputed Collection Samples																				
Group	All		JAWDHSC		MLOC		MMDSC		MSUFAL DSC		STAFS DSC		TXSTDSC		UTK DSC		WMed STARS			
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR		
AF	29	62.1	27	63	27	63	27	59.3	27	63	27	63	28	60.7	27	59.3	28	60.7		
AM	48	62.5	47	63.8	47	66	47	66	47	63.8	48	62.5	48	60.4	47	66	48	60.4		
BF	94	64.9	84	66.7	84	65.5	85	63.5	88	67	85	64.7	85	65.9	84	64.3	85	65.9		
BM	117	62.4	105	61.9	107	58.9	107	59.8	110	60.9	108	61.1	104	60.6	103	60.2	104	60.6		
CHM	77	48.1	73	46.6	77	46.8	73	50.7	73	47.9	73	49.3	73	46.6	72	48.6	73	46.6		
GTM	66	71.2	66	71.2	66	71.2	66	71.2	66	72.7	66	72.7	66	72.7	66	71.2	66	72.7		
HF	45	62.2	36	61.1	36	58.3	37	64.9	36	58.3	38	57.9	36	58.3	37	62.2	36	58.3		
HM	187	28.3	167	31.7	166	31.3	176	30.7	167	31.1	173	31.2	167	29.9	166	30.7	167	29.9		
JF	120	60.8	113	64.6	120	59.2	113	61.9	113	61.9	113	63.7	113	62.8	113	65.5	113	62.8		
JM	190	40.5	183	44.8	190	43.2	183	43.2	183	44.8	183	43.7	183	44.3	183	43.2	183	44.3		
VM	48	62.5	47	63.8	48	64.6	47	66	47	63.8	47	61.7	47	66	47	61.7	47	66		
WF	575	77.4	213	76.5	178	76.4	243	76.1	161	75.8	205	74.1	163	77.3	242	74.8	163	77.3		
WM	724	76.8	248	76.2	221	76	294	75.9	210	77.6	270	76.3	216	76.9	281	76.5	216	76.9		

Correct classification rates were also calculated for each individual within each documented collection, for the models using the total DSC sample combined with FD3 (Table 2.10) and models only using individuals from one collection combined with FD3 (Table 2.11). Collection classification accuracy rates are higher when all collections are included, for both non-imputed and imputed samples.

Table 2.10. Correct classification rates for each collection utilizing models combining Fordisc data with data from all eight DSCs.

Collection	Full FD3 Dataset		FD3 without MMDSC		FD3 without UTK DSC	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
<i>Collection Data Subsets with Complete Observations</i>						
JAWDHSC	37	78.4	37	78.4	37	78.4
MLOC	31	61.3	31	61.3	31	58.1
MMDSC	53	75.5	53	75.5	53	73.6
MSUFAL DSC	15	66.7	15	66.7	15	66.7
STAFS DSC	20	70.0	20	70.0	20	70.0
TXSTDSC	69	85.5	69	85.5	69	84.1
UTK DSC	29	75.9	29	75.9	29	75.9
WMed STARS	10	70.0	10	70.0	10	70.0
	264	72.9	264	72.9	264	72.1
<i>Collection Datasets with Missing Values Imputed</i>						
JAWDHSC	118	81.4	118	81.4	118	81.4
MLOC	76	56.6	76	56.6	76	56.6
MMDSC	207	71.0	207	71.0	207	72.5
MSUFAL DSC	35	60.0	35	60.0	35	60.0
STAFS DSC	144	75.0	144	75.0	144	75.7
TXSTDSC	239	79.5	239	79.5	239	79.5
UTK DSC	179	76.5	179	76.5	179	76.5
WMed STARS	37	70.3	37	70.3	37	73.0
	1035	71.3	1035	71.3	1035	71.9

Table 2.11. Correct classification rates for each collection utilizing models combining the Fordisc data with data only from individuals in that collection.

Collection	Full FD3 Dataset		FD3 without MMDSC		FD3 without UTK DSC	
	<i>n</i>	CCR	<i>n</i>	CCR	<i>n</i>	CCR
<i>Collection Data Subsets with Complete Observations</i>						
JAWDHSC	37	75.7	37	75.7	37	70.3
MLOC	31	58.1	31	58.1	31	54.8
MMDSC	53	71.7	53	73.6	53	73.6
MSUFAL DSC	15	66.7	15	66.7	15	66.7
STAFS DSC	19	68.4	19	68.4	19	63.2
TXSTDSC	69	84.1	69	84.1	69	84.1
UTK DSC	29	75.9	29	75.9	29	75.9
WMed STARS	10	70.0	10	70.0	10	70.0
	263	71.3	263	71.5	263	69.8
<i>Collection Datasets with Missing Values Imputed</i>						
JAWDHSC	118	79.7	118	79.7	118	76.3
MLOC	76	55.3	76	55.3	76	51.3
MMDSC	207	70.0	207	70.0	207	71.0
MSUFAL DSC	35	57.1	35	57.1	35	57.1
STAFS DSC	144	72.2	144	72.2	144	71.5
TXSTDSC	238	76.1	238	76.1	237	75.5
UTK DSC	179	76.0	179	76.0	179	73.2
WMed STARS	37	70.3	37	70.3	37	73.0
	1034	69.6	1034	69.6	1033	68.6

The outliers from the analyses were examined. Twenty-one individuals from the Fordisc sample were identified as outliers in every analysis. Of the 52 outliers identified throughout these analyses, 37 originated from the Fordisc sample and 15 from the DSC sample.

Figures 2.2 through 2.4 visualize the patterns of group separation at three levels of classification: source and population (see Figure 2.2); source and sex (see Figure 2.3); and, source, population, and sex (see Figure 2.4). These plots were examined to evaluate the morphological distinctions and overlaps between groups. Some separation between the Fordisc and DSC samples is observed in all three CVA plots, but the magnitude and degree of separation depends on the assessed grouping variable (i.e., source and sex or source and population). In Figure 2.2, six of the eight Fordisc populations are located in the lower right quadrant. Whereas in Figure 2.3, females and males are separated by the first canonical variate axis, while the female and male Forensic

Data Bank data are separated from all other collections along Can2.

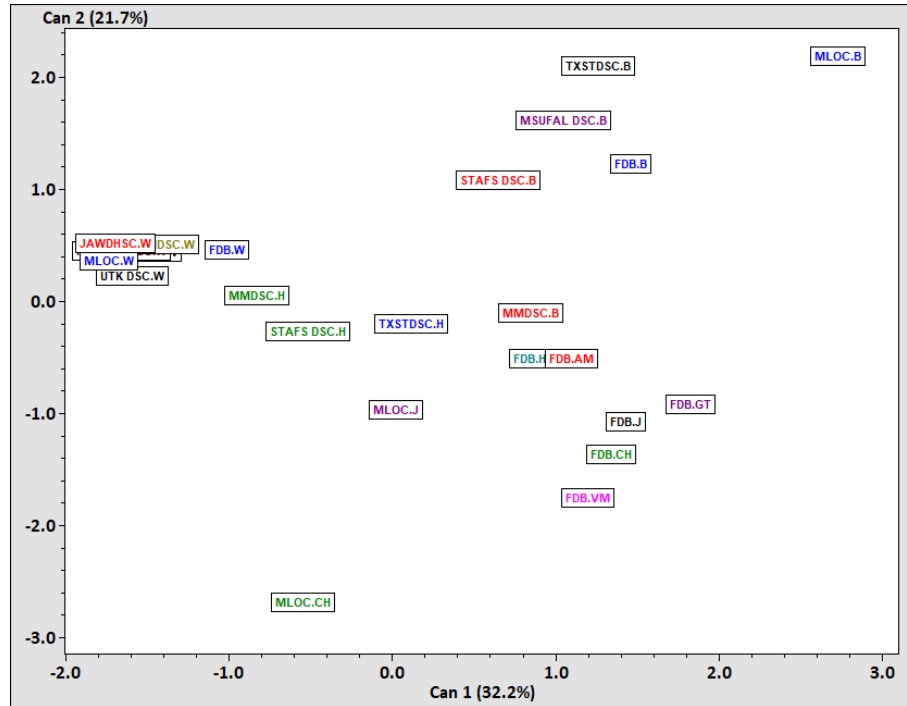


Figure 2.2. Canonical Variates Plot, by source and population groups. The plot displays the first two canonical variates (Can1 and Can2), which account for nearly 54% of the variance in the data. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005).

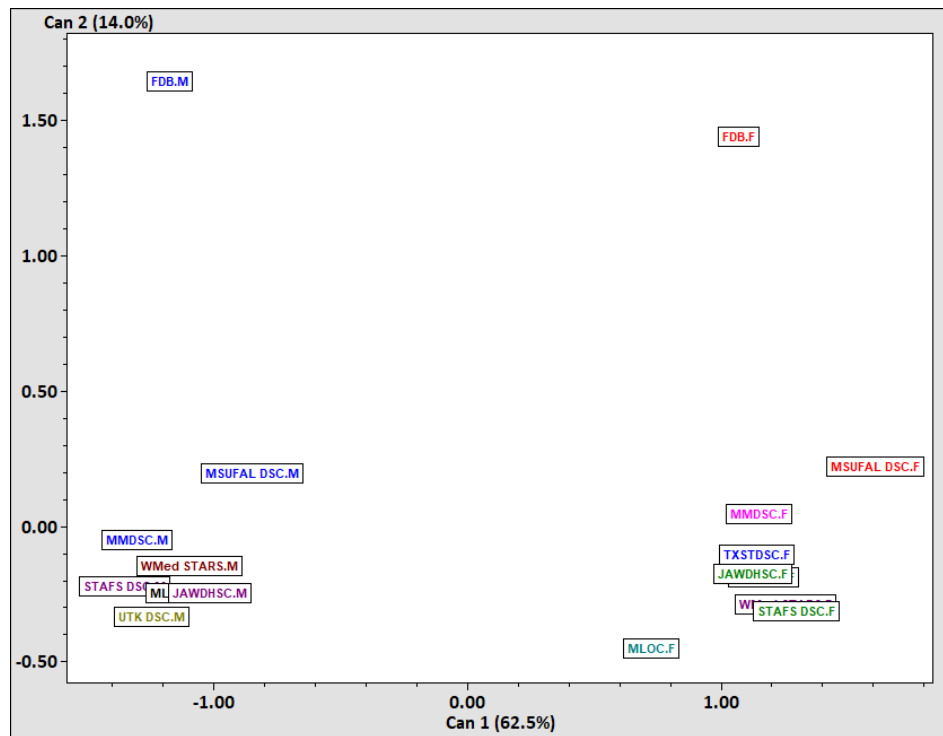


Figure 2.3. Canonical Variates Plot, by data source and sex. The plot displays the first two canonical variates (Can1 and Can2), which account for 76.5% of the model variance. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005).

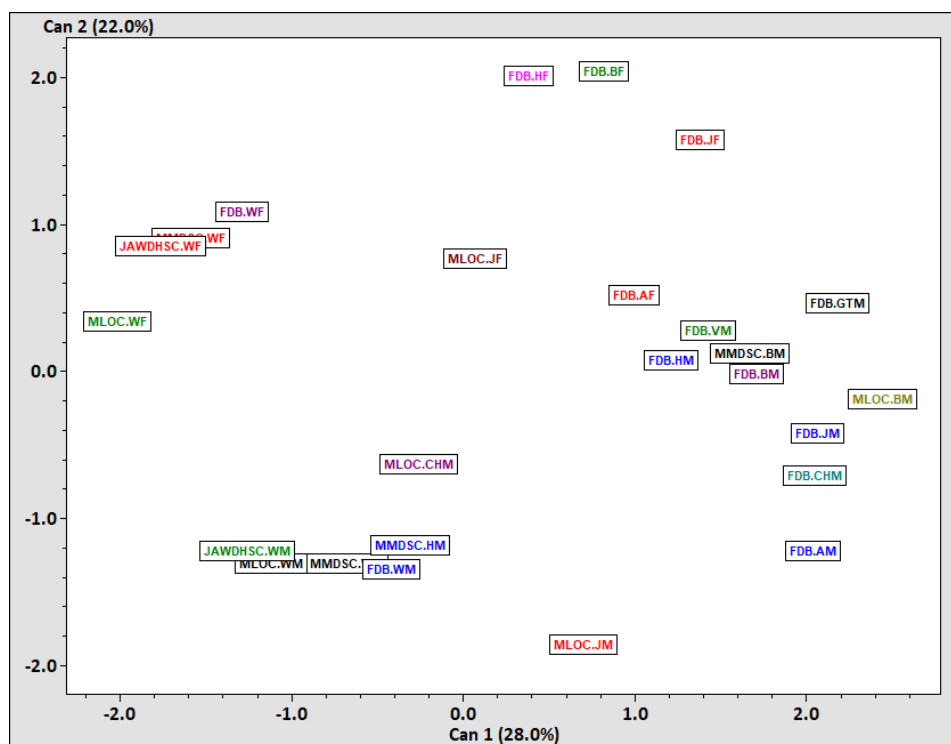


Figure 2.4. Canonical Variates Plot, by source, population, and sex. The plot displays the first two canonical variates (Can1 and Can2), which account for 50% of the model variance. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005).

The Mahalanobis distance (D^2) matrices calculated by Fordisc are illustrated in Figures 2.5 through 2.7 for the same three levels of classification: source and population (see Figure 2.5); source and sex (see Figure 2.6); and, source, population, and sex (see Figure 2.7). These dendrograms were examined to evaluate the similarities/dissimilarities between groups. The most similar groups include the European American Fordisc samples and the European American DSC samples (highlighted in Figures 2.5 and 2.7). However, when only source and sex are examined, males and females in documented collections are most dissimilar to males and females in Fordisc, respectively.

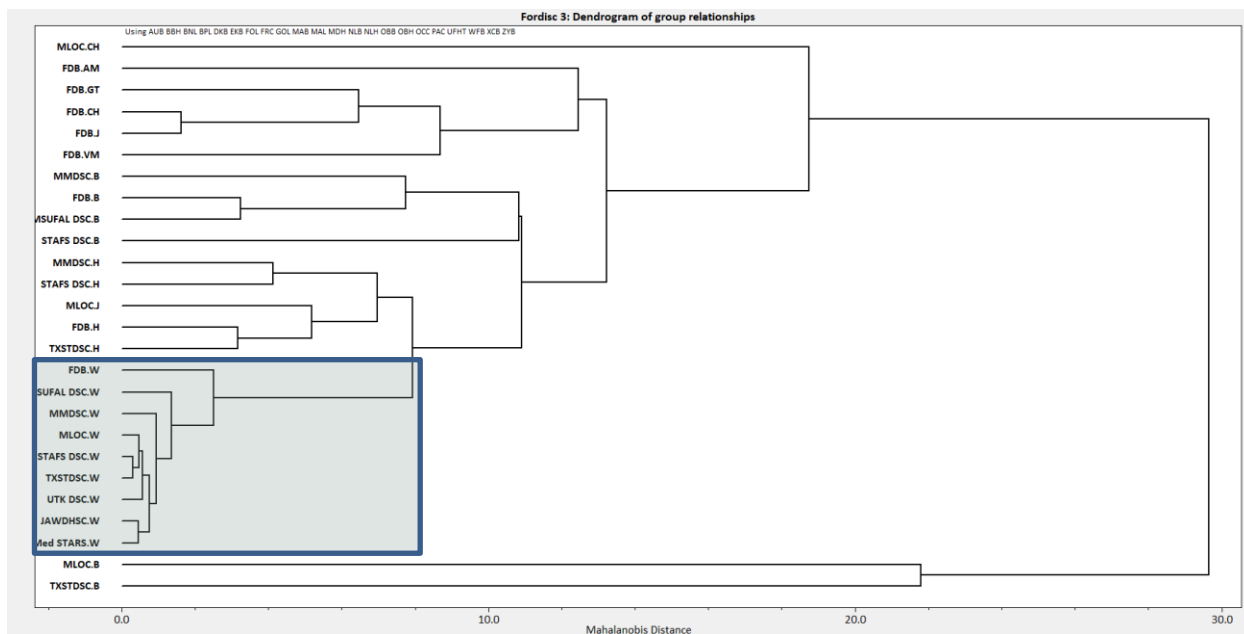


Figure 2.5. Hierarchical clustering dendrogram using the Mahalanobis distance (D^2) matrix calculated from source and population specific groups. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005). Clustering of European American samples are highlighted in green.

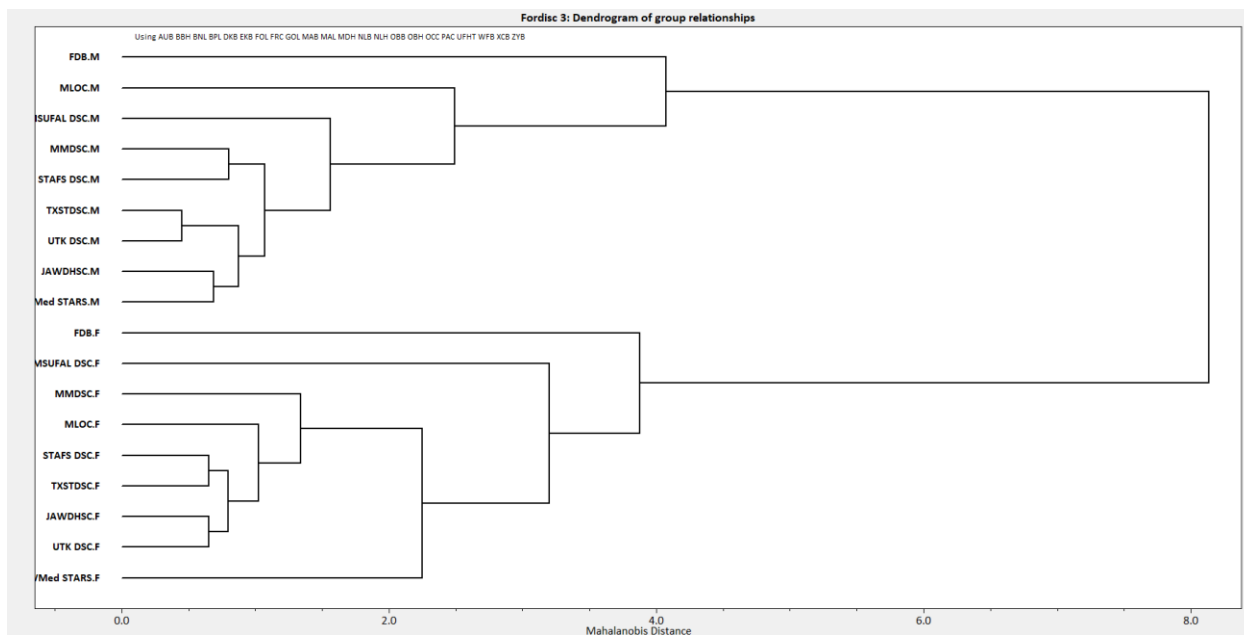


Figure 2.6. Hierarchical clustering dendrogram using the Mahalanobis distance (D^2) matrix calculated for source and sex specific groups. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005).

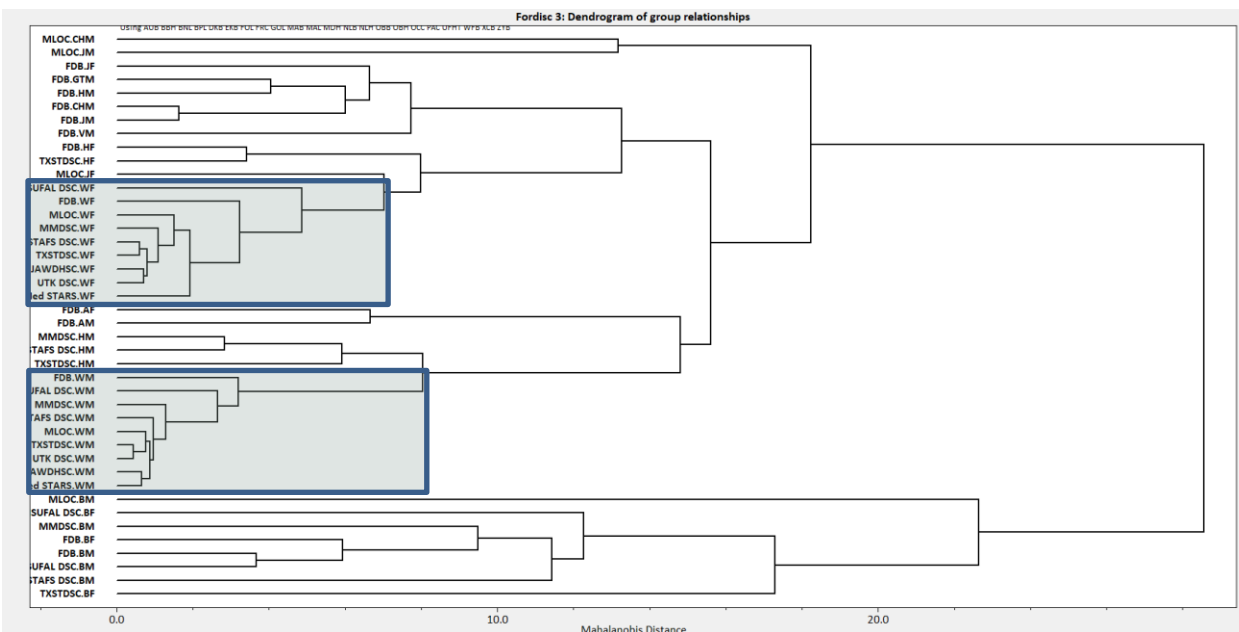


Figure 2.7. Hierarchical clustering dendrogram using the Mahalanobis distance (D^2) matrix calculated for source and population specific groups. This plot was downloaded directly from FD3, v. 3.1.322 (Jantz and Ousley 2005). Clustering of European American samples are highlighted in green.

DISCUSSION

This paper provides valuable insights into the representativeness of modern documented skeletal collections to the forensic case population. Overall, modern donated skeletal collection samples are not markedly different in cranial morphology from forensic cases (as represented by Fordisc), but these collections are only partially representative of those cases. Collections were most representative of the populations well-reflected in their samples, such as European Americans who comprise the largest population group in every collection (see Tables 2.3 and 2.6). The lack of diversity within donor populations hinders representation in skeletal collections. Of course, there has been marked improvement in representation compared to historic collections, which no longer accurately mirror cranial morphology of forensic cases even in populations well-reflected in collection samples (Jantz and Jantz 2000; Jantz and Meadows Jantz 2016; Jantz and Moore-Jansen 1987). More refined and nuanced population variability is essential for true

representativeness of collections to forensic cases.

Cranial morphology between each donated skeletal collection and the Fordisc samples does not differ substantially. Overall correct classification rates remained consistent or even improved with the addition of collection samples (see Table 2.8). Moreover, the 21 outliers identified in every analysis consistently originated from the Fordisc sample. If the documented skeletal collection samples were significantly different from the forensic cases, we would expect a greater proportion of the outliers from collections.

However, correct classification rates varied more when population/sex groups within each model are examined. Certain groups, such as Native American males, Hispanic males, and Japanese males and females, often had lower classification accuracies when the donated skeletal collection sample datasets were pooled with the Fordisc reference sample. In contrast, European Americans had higher classification accuracies in every model. An examination of these samples by source and population further highlights that disparity (see Figure 2.2); the European American sample in Fordisc clusters closely with European American samples from the donated skeletal collections, while the other population samples were separated and seemingly more dissimilar. This disparity improves when samples are classified by source, population, and sex (see Figure 2.4), however certain groups, such as Chinese males and Japanese males, continue to separate more than expected. This discrepancy may be attributed to smaller sample sizes and unbalanced sex distributions among non-European American groups in the documented skeletal collections, as smaller sample sizes provide less information for the model to learn from. Alternatively, this finding could suggest individuals from these groups are more poorly represented by the samples within these collections.

Correct classifications for individuals within documented skeletal collections were higher

when models included reference samples from other documented collections (see Tables 2.9 and 2.10). Individuals within collections seem more similar to other individuals in other DSCs than any DSC group is to the individuals in the Fordisc reference sample. This most likely represents the substantially larger samples of European American males and females in models when all collections are included—a fact clearly demonstrated when samples are classified by source and sex (see Figure 2.3). In those models, females and males in the Fordisc and DSC samples clearly separate and represent differences in population variability. These higher classification rates, particularly for European Americans, may be artificially inflating the overall classification accuracies of these models. Future research should use balanced population samples to assess the impact of this finding and assess whether other factors, such as differences in mean ages between collection and forensic data, could be contributing.

Three subsets of the Fordisc data (all FD3 individuals, FD3 without the MMDSC, and FD3 without the UTK DSC) were created to assess model performance with and without the donated skeletal collection samples. We hypothesized that the removal of the MMDSC and UTK DSC would only affect classification accuracies of the individuals currently housed in those two collections, since their removal should reduce the number of potentially similar individuals available in the reference data. However, neither collection was impacted by the removal of those samples from Fordisc (see Tables 2.10 and 2.11). Rather, examination of the overall population/sex classification accuracies indicate the removal of UTK DSC individuals from the Fordisc sample impacted **overall** model performance. All of the models excluding the UTK DSC individuals had lower overall correct classification accuracies (i.e., more error) than did any of the models where that sample was included.

The removal of UTK DSC individuals most likely impacted overall correct classification

because it reduced the reference sample size by 16% ($n = 248$). In contrast, removing the MMDSC only reduced the Fordisc sample by ~2% ($n = 30$). And this study used a subset of the total Fordisc sample; a thorough examination of the complete sample reveals the UTK DSC comprises nearly one-quarter (23%; $n = 574$) of the entire Fordisc reference data. Although J. Lawrence Angel is generally cited as the single largest contributor to the FDB (and thus FD3) reference sample (Jantz and Ousley 2005; Jantz 2019), the UTK DSC sample actually makes up the largest data from any one source. Thus, the influence of the UTK DSC on model performance cannot be understated.

Decreases in model performance after removal of the UTK DSC samples reveal inherent sample bias within the FD3 reference samples, but these biases may not impact the majority of forensic anthropological casework in the U.S. Still, the Fordisc reference data is largely influenced by one data source (UTK DSC), a documented skeletal collection rather than forensic case data. Other researchers have already demonstrated the need for robust and variable reference samples to maintain high classification accuracy (e.g., Birkby 1966; Fried et al. 2005; Go et al. 2019; Guyomarc'h and Bruzek 2011; Manthey et al. 2018). The Fordisc help file emphasizes how the program is not appropriate for the classification of archaeological populations or those belonging to demographic groups not available in reference samples, such as Asian Indian individuals (Jantz and Ousley 2005). A relatively straightforward fix for this dilemma: forensic anthropologists need to submit their forensic cases to the Forensic Anthropology Data Bank, regularly, to increase the diversity in sample sources, bolster reference sample sizes, and increase the number and variety of the demographic populations represented in Fordisc. This one action will improve estimations of population and sex for unknown decedents during forensic anthropological casework. Some bias is inherent in the use of Fordisc, as only identified forensic cases can be submitted to the FDB for inclusion within reference samples; however, continued case submission will mitigate sources of

bias related to inadequate representation.

CONCLUSION

This study found that modern documented skeletal collections are representative of forensic cases for population groups well reflected within those collections, such as European Americans. We suggest documented skeletal collections are valid sources of research in forensic anthropology for European Americans. Efforts to increase diversity in modern documented collections should increase their overall representativeness to forensic cases.

Our analysis also revealed the significant influence UTK DSC samples have on model accuracy in Fordisc. In order to increase model robustness and reliability in Fordisc, case submissions to the FDB are essential. This will ensure that reference samples represent cases across the United States and will reduce the influence that one data source currently has on Fordisc performance.

To further explore the representativeness of documented collections, future studies should conduct similar analyses with more balanced population and sex collection samples and assess how continued case submission to the FDB increases accuracy and performance in forensic anthropological casework.

BIBLIOGRAPHY

- Birkby, W. H. 1966. An evaluation of race and sex identification from cranial measurements. *American Journal of Physical Anthropology* 24:21-28.
- Campanacho, V., Ales Cardoso, F., & Ubelaker, D. H. 2021. Documented skeletal collections and their importance in forensic anthropology in the United States. *Forensic Sciences* 1:228-239. <https://doi.org/10.3390/forensicsci1030021>
- Christensen, A. M., & Passalacqua, N. V. 2018. *A laboratory manual for forensic anthropology*. London, UK: Academic Press. <https://doi.org/10.1016/C2016-0-03295-3>
- Davidson, M., & Morgan, R. 2022. A survey of ancestry estimation method preferences and utilization in forensic anthropology. Paper presented at the Proceedings of the 74th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA.
- Fleischman, J. M., & Crowder, C. M. 2019. Standard operation procedure for microscribe 3-dimensional digitizer and craniometric data.
- Fried D. L., Spradley, M. K., Jantz, R., & Ousley, S. 2005. The truth is out there: How not to use FORDISC. Paper presented at the Proceedings of the 74th Annual Meeting of the American Association of Physical Anthropologists, Milwaukee, WI.
- George, R. L., Zejdlik, K., Messer, D. L., & N. V. Passalacqua. 2022. The John A. Williams Human Skeletal Collection at Wester Carolina University. *Forensic Sciences* 2:362-370. <https://doi.org/10.3390/forensicsci2020026>
- Go, M. C., Jones, A. R., Algee-Hewitt, B. F. B., Dudzik, B., & Hughes, C. E. 2019. Classification trends among contemporary Filipino crania using Fordisc 3.1. *Forensic Anthropology* 2(4):1-11.
- Gocha, T. P., Mavroudas, S. R., & Wescott, D. J. 2022. The Texas State Donated Skeletal Collection at the Forensic Anthropology Center at Texas State. *Forensic Sciences* 2(1):7-19. <https://doi.org/10.3390/forensicsci2010002>
- Gordon, C. C., & Bradtmiller, B. (1992). Interobserver error in a large scale anthropometric survey. *American Journal of Human Biology*, 4(2): 253-263. <https://doi.org/10.1002/ajhb.1310040210>
- Guyomarc'h, P., & Bruzek, J. 2018. Accuracy and reliability in sex determination from skulls: A comparison of Fordisc 3.0 and the discriminant function analysis. *Forensic Science International* 208(1-3): 180.e1-e6. <https://doi.org/10.1016/j.forsciint.2011.03.011>
- Herrera, B., Hanihara, T., & Godde, K. 2014. Comparability of multiple data types from the Bering strait region: Cranial and dental metrics and nonmetrics, mtDNA, and Y-chromosome DNA. *American Journal of Biological Anthropology* 154(3):334-348. <https://doi.org/10.1002/ajpa.22513>

- Hughes, C. E., & Juarez, C. A. 2018. Learning from Our Casework: The Forensic Anthropology Database for Assessing Methods Accuracy (FADAMA). NIJ 2018-DU-BX-0213.
- Jantz, R. L. 2019. Title discoveries from the Forensic Anthropology Data Base: Modern American skeletal change & the case of Amelia Earhart. *The FASEB Journal*, 33:202.1-202.1. https://doi.org/10.1096/fasebj.2019.33.1_supplement.202.1
- Jantz, L. M., & Jantz, R. L. 2000. Secular change in cranial facial morphology. *American Journal of Human Biology*, 12(3), 327-338. [https://doi.org/10.1002/\(SICI\)1520-6300\(200005/06\)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1520-6300(200005/06)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1)
- Jantz, R. L., & Meadows Jantz, L. 2016. The remarkable change in Euro-American cranial shape and size. *Human Biology* 88(1): 56-64.
- Jantz R. L., & Moore-Jansen, P. H. 1987. *Final report to the National Institute of Justice: Grant No. 85-IJ-CX-0021. Department of Anthropology*, University of Tennessee, Knoxville.
- Jantz, R. L., & Ousley, S. D. 2005. FORDISC 3: Computerized Forensic Discriminant Functions. Version 3.1. The University of Tennessee, Knoxville.
- Jantz, R. L., & Ousley, S. D. 2012. Introduction to Fordisc 3. In: N. R. Shirley & M. A. Tersigni-Tarrant, eds. *Forensic anthropology: An Introduction*. Boca Raton, FL: CRC Press. pp. 253-269. <https://doi.org/10.1201/b12920>
- Kachigan, S. M. 1982. *Multivariate statistical analysis: A conceptual introduction, 2nd edition*. Radius Press.
- Komar, D., & Grivas, D. 2008. Manufactured populations. What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum Documented Collection. *American Journal of Biological Anthropology* 137(2): 244-233. <https://doi.org/10.1002/ajpa.20858>
- Mann, R. W., Labrash, S., & Lozanoff, S. 2020. A new osteological resource at the John A. Burn's School of Medicine. *Hawai'i Journal of Health & Social Welfare* 79(6): 202-203.
- Manthey, L., Jantz, R. L., Vitale, A., & Cattaneo, C. 2018. Population specific data improves Fordisc's performance in Italians. *Forensic Science International* 292: 263.e1-e7. <https://doi.org/10.1016/j.forsciint.2018.09.023>
- Ousley SD. 2014. *3Skull version 1.76*, <https://www.statemachine.net/software/3Skull/>
- Ousley, S. D., & Jantz, R. L. 2012. Fordisc 3 and Statistical Methods for Estimating Sex and Ancestry. In: Dennis C. Dirkmaat, ed. *A Companion to Forensic Anthropology*. London, UK: CRC Press. pp. 311-329.

- R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>
- Relethford, J. H. 1994. Craniometric variation among modern human populations. *American Journal of Physical Anthropology* 95:53-62.
- Relethford, J. H. 2009. Race and global patterns of phenotypic variation. *American Journal of Biological Anthropology*, 139(1): 16-22. <https://doi.org/10.1002/ajpa.20900>
- Roseman, C. C. 2004. Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *PNAS* 101(35):12824-12829.
- Roseman, C. C., & Weaver, T. D. 2004. Multivariate apportionment of global craniometric diversity. *American Journal of Physical Anthropology* 125:257-263.
- Sparks, C. S., & Jantz, R. L. 2002. A reassessment of human cranial plasticity. *PNAS*, 99(23): 14636-14639. <https://doi.org/10.1073/pnas.222389599>
- STAFS Research. 2024. Southeast Texas Applied Forensic Science Facility, Sam Houston State University. Retrieved October 13, 2024, from <https://ifrti.org/STAFS/research.html>
- Utermohle, C. J., & Zegura, S. L. (1982). Intra- and interobserver error in craniometry: A cautionary tale. *American Journal of Biological Anthropology*, 57(3): 303-310. <https://doi.org/10.1002/ajpa.1330570307>
- Utermohle, C. J., Zegura, S. L., & Heathcote, G. M. (1983). Multiple observers, humidity, and choice of precision statistics: Factors influencing craniometric data quality. *American Journal of Biological Anthropology*, 61(1): 85-95. <http://dx.doi.org/10.1002/ajpa.1330610109>
- Winburn, A. P., Jennings, A. L., Steadman, D. W., & DiGangi, E. A. 2022. Ancestral diversity in skeletal collections: Perspectives on African American body donation. *Forensic Anthropology* 5(2):141-152. <https://doi.org/10.5744/fa.2020.1023>

MANUSCRIPT 3. GEOSTATISTICAL AND SPATIAL ANALYSIS OF SELECTION BIAS IN DOCUMENTED SKELETAL COLLECTIONS

INTRODUCTION

Sampling is an essential part of the research process, dictating and defining how representative a sample is to the population under investigation (Smith and Noble 2014; Smith 2019). In biological anthropology, many studies utilize samples from documented skeletal collections, which contain hundreds, or even thousands, of human skeletal remains from known individuals. These collections ensure an adequate sample is available for research purposes, which is imperative for permitting unbiased sampling strategies. However, because so many researchers use the same documented skeletal collections for their research, the extent to which these samples are truly random, and aspects of selection bias inherent within each collection, need to be documented and fully understood.

Various factors influence sample selection, including time constraints, the requirements of a study (e.g., balanced sex samples), a collection's layout, and the level of involvement of the curator or collection manager. For example, when time issues arise, researchers may forego individuals on higher shelves that necessitate cumbersome ladders. Of course, if those individuals are imperative to meet the demands of a study, the researcher would take the time to seek out that individual, even on the top shelves. Consider also that every researcher does not employ simple random sampling techniques. There are research questions that cannot be answered without adequate reference data from multiple groups. However, randomness during sampling can avoid over or under sampling individuals in a collection simply because of their location within the storage area.

Geostatistics and spatial analysis include tools to directly evaluate the spatial distribution

of individuals within a collection and how that distribution influences their likelihood of being sampled for research. Originally developed by geologists to predict deposit locations at unsampled sites, geostatistics have since been applied to biological distance studies (*c.f.* Relethford 2008; Hefner 2017). These approaches use variogram analysis to quantify the relationship between a biological variable, such as sex, and the physical locations of individuals. Variogram analysis identifies the magnitude, extent, and pattern of any spatial correlation. The identified (and quantified) correlation is used for kriging, an interpolation method to estimate unknown values between known individuals based on nearby observations and the associated spatial autocorrelation identified during the variogram analysis. Importantly, kriging does not imply the presence of additional individuals; rather, by estimating these new data points, kriging generates a smoothed contour plot for visualization and interpretation of the underlying spatial patterns (Borcard et al. 2018; Legendre and Legendre 1998; Relethford 2008).

Using these methods of analysis in tandem, we can identify whether a relationship exists between sampling frequencies and the physical location of individuals within a collection. Likewise, demographic data for the individuals within a collection can be used in similar analyses to identify additional patterns (e.g., more females at one end of a shelf). By identifying patterns explained by demographic or temporal factors, we can then parse out other patterns having the potential to impact sampling strategies.

This research aims to investigate whether sampling selection bias exists in biological anthropological research using documented skeletal collections. To identify selection bias, we use demographic data, sampling frequency data, and the physical layout information for five United States (U.S.)-based documented skeletal collections.

MATERIALS AND METHODS

Five U.S.-based documented skeletal collections are examined in this study. Two historic collections—the Hamann-Todd Human Osteological Collection (HTH Collection) and the Robert J. Terry Anatomical Collection (Terry Collection)—were chosen as foundational collections in biological anthropology research. The University of Tennessee, Knoxville, Donated Skeletal Collection (UTK DSC), the Maxwell Museum’s Documented Skeletal Collection (MMDSC), and the Texas State University Donated Skeletal Collection (TSXTDSC) were included as modern, well-researched collections. Background information for each collection and their physical space layout are provided below.

Hamann-Todd Human Osteological Collection (HTH Collection)—Originally established at Case Western Reserve University in Cleveland, Ohio, the HTH Collection comprises over 3,000 individuals collected between 1893 and 1938. The HTH Collection is now located at the Cleveland Museum of Natural History (CMNH) in Cleveland, Ohio (Muller et al. 2017). Unlike many other documented collections, the skeletal remains of all individuals in the HTH Collection are stored between two rooms: skulls in one room, postcranial skeletons in another. In both rooms, the individuals are stored in high capacity rolling compactor shelving units. Only one row of shelves may be accessed at any time. Skulls are stored in boxes 13 rows high with 4 boxes in each row (Figure 3.1A); postcranial skeletons are stored within large, flat drawers 19 rows high with 1 drawer per row (see Figure 3.1B). The museum provides a ladder for access to higher shelves, but the ladder must be removed from the row when not in use to ensure other researchers can access the collection. Individuals in the HTH Collection are stored chronologically, from left to right. Of note, the CMNH halted research access to the HTH Collection in 2023 for major renovations, which will change the collection layout moving

forward (Collections 2024); however, the physical layout discussed herein has been in place since the compactor shelves were initially installed in the 1980s (Anonymous interviewee A, personal communication, 2024).

A	B	C	D	E	F	G	H
	1						
	2						
	3						
	4						
	5						
	6	1					
1 2 3 4	7	2	1 2 3 4				
5 6 7 8	8	3	5 6 7 8				
9 10 11 12	9	4	9 10 11 12				
13 14 15 16	10	5	13 14 15 16	1	1 2 3 4		
17 18 19 20	11	6	17 18 19 20	2	5 6 7 8	1 4 7 10 13 16 19 22	1 2 3 4 5 6 7
21 22 23 24	12	7	21 22 23 24	3	9 10 11 12	2 5 8 11 14 17 20 23	8 9 10 11 12 13 14
25 26 27 28	13	8	25 26 27 28	4	13 14 15 16	3 6 9 12 15 18 21 24	15 16 17 18 19 20 21
29 30 31 32	14	9	29 30 31 32	5	17 18 19 20	25 28 31 34 37 40 43 46	22 23 24 25 26 27 28
33 34 35 36	15	10	33 34 35 36	6	21 22 23 24	26 29 32 35 38 41 44 47	29 30 31 32 33 34 35
37 38 39 40	16	11	37 38 39 40	7	25 26 27 28	27 30 33 36 39 42 45 48	36 37 38 39 40 41 42
41 42 43 44	17	12	41 42 43 44	8	29 30 31 32	49 52 55 58 61 64 67 70	43 44 45 46 47 48 49
45 46 47 48	18	13	45 46 47 48	9	33 34 35 36	50 53 56 59 62 65 68 71	50 51 52 53 54 55 56
49 50 51 52	19	14	49 50 51 52	10	37 38 39 40	51 54 57 60 63 66 69 72	57 58 59 60 61 62 63

Figure 3.1. Representation of how individuals are chronologically ordered within one cabinet in each collection: A) HTH Collection cranial cabinets; B) HTH Collection postcranial cabinets; C) tall cabinets in Terry Collection; D) short cabinets in Terry Collection; E) old UTK DSC cabinets; F) current UTK DSC cabinets; G) MMDSC cabinets; and, H) TXSTDSC cabinets. In cabinets with single rows of individuals (B, C, and E), storage continues at the top of the next cabinet, located to the right, with the same ordering (i.e., top to bottom).

Robert J. Terry Anatomical Collection (Terry Collection)— Originally established at Washington University in St. Louis, Missouri, the Terry Collection comprises over 1,700 individuals accessioned between 1910 and 1967. The Terry Collection is now housed within the Smithsonian Institution, National Museum of Natural History, in Washington, D.C. (Hunt and Albanese 2005). This collection has changed physical layouts multiple times and a current moratorium on research, established in 2023 (Sholts 2024), restricts study of the current layout. This study focuses on the historic layout used from circa 1980 to 2005 (Anonymous interviewee

B, personal communication, 2024). Individuals were stored in drawers in three different types of cabinets. Fifteen cabinets were located in an analytical room; the remainder were situated outside of that room, along the wall in a long hallway (Anonymous interviewee B, personal communication, 2024). Each individual was stored in a drawer, in cabinets either 13 or 14 rows high and stored from left to right within each cabinet; individuals in tall cabinets are ordered 1 per row, while individuals in shorter cabinets are ordered 4 per row (see Figure 3.1C-D). Wheeled step ladders were available to researchers. Unlike many other documented collections, accession numbers at the Terry Collection were reused if a particular individual was de-accessioned. This complicated the numbering system, particularly when approximately 90 individuals were added back into the collection by Mildred Trotter (Hunt and Albanese 2005). To distinguish these cases, accession numbers were appended with “R”. For example, 15, 15R, 15RR, and 15RRR represent four females with the same base accession number (Hunt and Albanese 2005). Individuals are stored chronologically with the exception of individuals with multiple “R’s” (e.g., 15RR, 15RRR), as well as any individuals added back into the collection (e.g., 15)—these individuals are organized chronologically at the very end of the collection. Individuals with a single “R” (e.g., 15R) are stored in the original chronological position within the collection. Storage of individuals in the collection begins with the cabinets in the hallway, then in the analytical room, and concludes with a series of additional cabinets in the hallway.

University of Tennessee, Knoxville, Donated Skeletal Collection (UTK DSC)—Located on the main campus of the University of Tennessee, Knoxville, in Knoxville, Tennessee, the UTK DSC was established in 1981 and currently contains over 1,800 donations (Body Donation 2024). The collection, which continues to accept donations today, changed physical locations in 2017 (Anonymous interviewee C, personal communication, 2024); both physical layouts are

examined in this study. In each one, individuals are stored in archive-quality cardboard boxes. In the pre-2017 physical layout, these boxes were stored on standalone, wire rack cabinets, five shelves high, with two stacked cardboard boxes per shelf. Most individuals required one box; however, infrequently two boxes were necessary. A step stool was provided for researchers to access higher shelves. In 2017, the collection was relocated to a new space incorporating high capacity rolling compactor shelving units. The same boxes are used, but the new shelves are ten levels high, and the boxes are no longer stacked. Access to the top shelves does require a ladder, which is provided to visiting researchers; collection protocols oblige a second person in the room when the ladder is in use (Forensic Anthropology Center 2022). Today, the UTK DSC skeletal remains are stored in one room, and a separate analytical room is nearby. Visiting researchers are permitted up to nine boxes in the analytical room at any one time. In both layouts, individuals are organized chronologically, from left to right across each cabinet; in the old layout, individuals were ordered 1 per row, while in current layout, cabinets have 4 boxes per row (see Figure 3.1E-F).

Maxwell Museum's Documented Skeletal Collection (MMDSC)—Located in the Maxwell Museum on the University of New Mexico's campus in Albuquerque, New Mexico, the MMDSC contains over 300 individuals. The MMDSC still actively accepts donations (Komar and Grivas 2008). Individuals are stored in archive-quality cardboard boxes, with one to two boxes per person, on standalone cabinets, three levels high. Each level accommodates up to three vertically stacked boxes; as such, a ladder is not necessary to access the top shelf. Analytical space is available adjacent to the cabinets and downstairs from the collection. When using the downstairs analytical space, boxes must be transferred via a walled lift (e.g., dumbwaiter). The MMDSC is currently (2024) upgrading to larger boxes that can hold all skeletal elements from

the same individual (Anonymous interviewee D, personal communication, 2024). Currently, remains are primarily stored in chronological order, progressing from left to right, with each row accommodating 24 boxes (see Figure 3.1G). Exceptions include multiple subadult remains housed in the same box and larger boxes stored separately for ease.

Texas State Donated Skeletal Collection (TXSTDSC)—Located in the Forensic Anthropology Center at Texas State University, San Marcos, in San Marcos, Texas, the TXSTDSC contains over 700 donations and is still accepting donations (Gocha et al. 2022). The TXSTDSC collection moved to its current location in 2018 (Anonymous interviewee E, personal communication, 2024). This study focuses on the current physical layout. Donated remains are stored in large archive-quality cardboard boxes, typically one box per individual. These boxes are placed on tall standalone wire racks, each having nine shelves. A step stool is provided to access individuals on the top shelves. Individuals are primarily organized chronologically, progressing from left to right, with 7 boxes per row (see Figure 3.1H), although exceptions for individuals requiring larger boxes exist. The collection cabinets are arranged in an “L” shape adjacent to the analytical space designated for Texas State University students and visiting researchers.

The first author (RRD) examined all current collection layouts; however, the historic layouts for the Terry Collection and UTK DSC were recreated using verbal descriptions, historic images, and anonymous interviews. While they are close approximates, these two physical layouts may not perfectly capture the complete layout.

Sample strategies and collection use were assessed using anonymous survey data sent to volunteer participants, 18 years or older, using the University of Florida Forensic Anthropology Listserv. This listserv is available to anyone with an interest in forensic anthropology, including

students and professionals. The survey asked participants to provide the accession numbers of individuals used in their research sample, as well as some basic information about their project (e.g., year, general focus; see *Appendix A*). To supplement the survey responses, master's theses and doctoral dissertations obtained from ProQuest were examined for sample tables and sampling strategy information; those listing sample accession numbers were culled, collected, and utilized, along with the study's focus and year of publication.

Demographic variables for individuals within each collection were required to contextualize any detected patterns. Age, population affinity, and biological sex data were provided by collection managers for all individuals in the HTH Collection and nearly all individuals in the MMDSC. These data were obtained for individuals housed in the other three collections through the aforementioned sample tables in master's theses and doctoral dissertations.

To further contextualize sampling patterns and strategies, anonymous interviews were conducted with individuals familiar with these documented skeletal collections to better understand restrictions on collection access or collection sampling, collection management, changes to collection layouts over time, and the assignment of new accession numbers (see *Appendix B*).

Geostatistical and Spatial Analysis

To capture the physical spatial layout of individuals within each collection, XY coordinates were assigned to each individual within each cabinet/box/shelf. The x-coordinate indicates the horizontal position across the cabinets and the y-coordinate corresponds to the vertical level. For example, in the old UTK DSC layout, the first individual is located on the first cabinet at coordinates (1,10), representing the leftmost position on the 10th vertical shelf. The

second individual is at coordinate (1,9), the third (1,8), etc. (Figure 3.2). All x-coordinates begin with the first documented individual in a collection and extend to the last, based on the sequential order provided by the curator and accession numbers. For ease, this approach treats the physical space as a single, continuous horizontal plane, thereby creating a two-dimensional layout suitable for spatial analysis. To preserve the spatial context within and between cabinets, cabinet numbers were also recorded for each individual.

Vertical Shelf Height	Cabinet 1					...	Cabinet N				
	1 , 10	2 , 10	3 , 10	4 , 10	5 , 10		Z-4 , 10	Z-3 , 10	Z-2 , 10	Z-1 , 10	Z , 10
	1 , 9	2 , 9	3 , 9	4 , 9	5 , 9		Z-4 , 9	Z-3 , 9	Z-2 , 9	Z-1 , 9	Z , 9
	1 , 8	2 , 8	3 , 8	4 , 8	5 , 8		Z-4 , 8	Z-3 , 8	Z-2 , 8	Z-1 , 8	Z , 8
	1 , 7	2 , 7	3 , 7	4 , 7	5 , 7		Z-4 , 7	Z-3 , 7	Z-2 , 7	Z-1 , 7	Z , 7
	1 , 6	2 , 6	3 , 6	4 , 6	5 , 6		Z-4 , 6	Z-3 , 6	Z-2 , 6	Z-1 , 6	Z , 6
	1 , 5	2 , 5	3 , 5	4 , 5	5 , 5		Z-4 , 5	Z-3 , 5	Z-2 , 5	Z-1 , 5	Z , 5
	1 , 4	2 , 4	3 , 4	4 , 4	5 , 4		Z-4 , 4	Z-3 , 4	Z-2 , 4	Z-1 , 4	Z , 4
	1 , 3	2 , 3	3 , 3	4 , 3	5 , 3		Z-4 , 3	Z-3 , 3	Z-2 , 3	Z-1 , 3	Z , 3
	1 , 2	2 , 2	3 , 2	4 , 2	5 , 2		Z-4 , 2	Z-3 , 2	Z-2 , 2	Z-1 , 2	Z , 2
	1 , 1	2 , 1	3 , 1	4 , 1	5 , 1		Z-4 , 1	Z-3 , 1	Z-2 , 1	Z-1 , 1	Z , 1

Figure 3.2. Example of coordinates given according to location on cabinets. X coordinates span from 1 to Z on Cabinets 1 to N, while Y coordinates span from 1 to 10.

Using the XY coordinate data for individuals within each collection, geostatistical and spatial analyses were conducted in R (R Core Team 2024) using the “sp”, “gstat”, and “automap” packages (Hiemstra et al. 2009; Pebesma 2004; Pebesma et al. 2005). Open-source code is available from the authors.

Geostatistical methods use known XY coordinate data to assess a third variable, Z, over geographic space (Borcard et al. 2011; Legendre and Legendre 1998; Relethford 2008). First, a measure of dissimilarity is calculated between known Z-values, producing an experimental variogram to quantify spatial autocorrelation within the dataset (Oliver and Webster 2015). A theoretical variogram model is then fitted to the experimental variogram to characterize that spatial relationship. Straight, non-undulating models indicate no spatial correlation, while sloping or undulating models suggest a relationship between geographic distance and the

parameter of interest (Legendre and Legendre 1998).

Model parameters describe the spatial structure (Oliver and Webster 2015). The sill indicates the total variance explained by geographic distance, while the range indicates the distance over which that spatial autocorrelation exists. The partial sill (*psill* or C_1) is the proportion of the variance explained by the spatial autocorrelation, while the nugget effect (C_0) is the variance at zero distance. The relative nugget effect ($C_0/(C_0 + C_1)$) indicates the proportion of variance attributable to noise, such as measurement error, sampling error, or variability at smaller scales than those addressed by the sampling interval (Legendre and Legendre 1998). A low relative nugget effect indicates the data are well explained by spatial structure.

Following variogram analysis, kriging is used to interpolate unknown values of Z across space (Oliver and Webster 2015). This regression-based method calculates values using weighted averages of known Z-values, where the weights are assigned as a measure of the spatial autocorrelation identified during variogram analysis (Relethford 2008). Using those interpolated values, smoothed contour maps visualize the underlying spatial relationship.

Variogram analysis and kriging were first conducted on sampling frequency distribution data to evaluate the spatial distribution of sampled individuals across each collection. This analysis identified spatial patterns related to sampling density, including the range, magnitude, and extent of spatial autocorrelation. The same geostatistical approach was then applied to age, population, and sex to assess whether they exhibited spatial clustering or autocorrelation within each collection, and, if so, how those patterns may explain sampling frequency. Variograms were generated to quantify these spatial relationships; kriging visualized these patterns.

After completing the geostatistical and spatial analyses, a series of t-tests, one-way analyses of variance (ANOVAs), and two-way ANOVAs were conducted. These assessed the

impact of demographic factors (e.g., age, sex), spatial variables (e.g., shelf height), and temporal factors on sampling frequencies.

RESULTS

Survey and Data Compilation Results

Seventy study samples were available for analysis (Table 3.1). The survey received sixteen responses. Forty-seven master's theses and dissertations that included detailed sample tables with individual accession numbers, by collection were identified through ProQuest. Several of the studies used individuals from multiple collections, so 54 total collection samples were identified.

Table 3.1. Breakdown of samples used from each collection by ProQuest studies and submitted studies through the survey.

Collection	Survey Samples (n)	ProQuest Samples (n)	Total Samples
HTH Collection – Crania Layout	3	5	8
HTH Collection – Postcranial Layout	3	7	10
Terry Collection	3	5	8
UTK DSC – Old Layout	2	20	22
UTK DSC – Curren Layout	2	5	7
MMDSC	2	3	5
TXSTDSC	1	9	10
Total	16	54	70

Donated skeletal collection demographic data was culled from the master's theses and dissertations available through ProQuest, supplementing existing data for the HTH Collection and MMDSC (Table 3.2). Age data was the most limited factor, ranging from 47.8% to 99.6% of the individuals within collections. Sex data was the most complete (61.4% to 100%). Missing data were primarily associated with individuals more recently added to collections, since some of the studies were conducted before those individuals were present in the collection.

Table 3.2. Breakdown of demographic data available for each collection. Total sample counts include individuals already stored within the collection and do not account for individuals currently being processed.

Collection	Total Sample (<i>n</i>)	Age Data		Sex Data		Population Data	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
HTH Collection – Cranial Layout	2,840	2,829	99.6	2,838	99.9	2,840	100
HTH Collection – Postcranial Layout	2,971	2,959	99.6	2,970	100	2,971	100
Terry Collection	1,728	826	47.8	1,061	61.4	1,045	60.5
UTK DSC – Old Layout	1,260	932	74.0	996	79.0	974	77.3
UTK DSC – Current Layout	1,726	1,110	64.3	1,174	68.0	1,152	66.7
MMDSC	313	277	88.5	279	89.1	269	85.9
TXSTDSC	588	475	80.8	478	81.3	480	81.6

HTH Collection – Cranial Layout

Of the eight studies we identified using cranial samples from the HTH Collection, 861 accession numbers were available for analysis. In total, 527 individuals (18.6%) were sampled (Figure 3.3B).

The geospatial analysis of the cranial material (as a sampling frequency) at the HTH Collection is presented in Figure 3.3. The experimental variogram model contains a substantial relative nugget effect (*nugget effect* = 0.48; *psill* = 0.09), accounting for 83.5% of the total variance.

The fitted spherical variogram model has a range of 46.09 cabinets, indicating spatial autocorrelation persists up to approximately 46 cabinets, beyond that the data are considered spatially independent (Figure 3.3A). Kriging interpolation identified one large area of higher sampling density, as well as two smaller areas near the middle of the collection (Figure 3.3C).

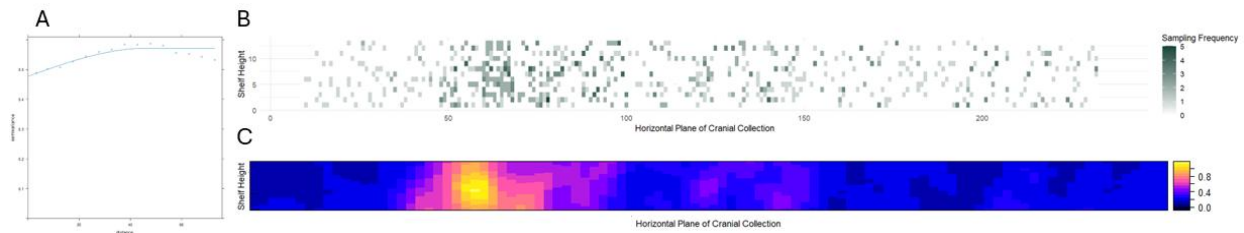


Figure 3.3. Spatial analysis of skull sampling from the HTH Collection. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the HTH cranial shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by yellow.

Results of the t-test and one-way ANOVAs are provided in Table 3.3. Females are sampled at a higher frequency than males, across all population groups (Figure 3.4). Shelf height does not impact sampling frequency, but cabinet number does—individuals in cabinets 3, 4, and 5 are sampled at a higher rate than those at either end of the collection. Examination of population distributions across the collection indicate these cabinets contain higher concentrations of African American individuals (Figure 3.5). In fact, the single most sampled individual ($n = 5$) was an African American female located in cabinet 4. Sampling outside of these cabinets (indicated by purple in the center of Figure 3.3C) seems to be associated with higher concentrations of Asian American and Native American individuals (see Figure 3.5) and/or females (Figure 3.6).

Table 3.3. Results of *t*-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency for cranial samples at the HTH Collection.

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = 23.33	3,678	<0.01	N/A	
Population	ANOVA	F = 29.66	3	<0.01	African American v. European American	<0.01
Population and Sex	ANOVA	F = 99.63	6	<0.01	African American females v. African American males, European American females, and European American males	<0.01
					European American males v. African American males and European American females	<0.01
Age (Decade)	ANOVA	F = 16	10	<0.01	0-10 v. 11-19, 20-29, 30-39, and 40-49	<0.01
					0-10 v. 50-59 and 80-89	<0.05
					10-19 v. 40-49	<0.05
					10-19 v. 50-59, 60-69, and 70-79	<0.01
					20-29 v. 30-39, 40-49, 50-59, 60-69, and 70-79	<0.01
					20-29 v. 80-89	<0.05
					30-39 v. 40-49 and 50-59	<0.01
					30-39 v. 70-79	<0.05
					40-49 v. 60-69	<0.05
Shelf Height (1-13)	ANOVA	F = 0.671	12	0.78	N/A	
Cabinet Number (1-12)	ANOVA	F = 19.86	11	<0.01	Cabinet 3 v. Cabinet 1, Cabinet 2, Cabinet 4, Cabinet 9, Cabinet 10, and Cabinet 12	<0.01
					Cabinet 3 v. Cabinet 6, Cabinet 8, and Cabinet 11	<0.05
					Cabinet 4 v. Cabinet 1, Cabinet 2, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, Cabinet 9, Cabinet 10, Cabinet 11, and Cabinet 12	<0.01
					Cabinet 5 v. Cabinet 1, Cabinet 2, Cabinet 6, Cabinet 8, Cabinet 9, Cabinet 10, Cabinet 11, and Cabinet 12	<0.01

*Only significant pairwise comparisons included for significant ANOVAs.

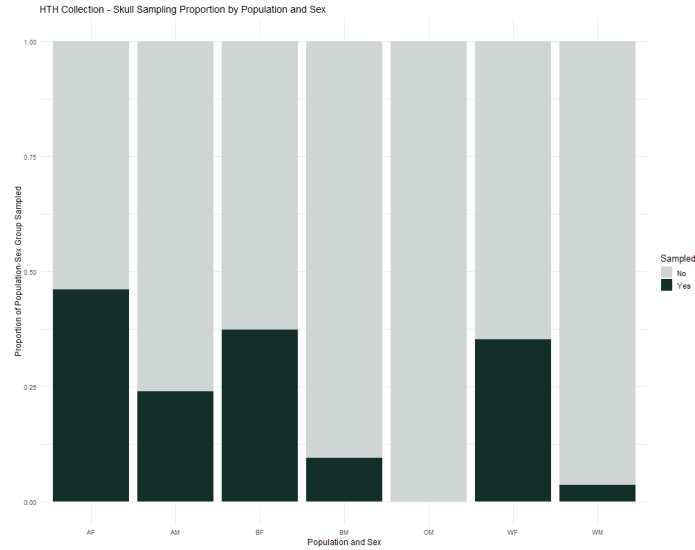


Figure 3.4. Proportion of individuals sampled by population and sex group. “M” and “F” = male and female. “A”, “B”, “O”, and “W” = Native American, African American, Asian American, and European American, respectively.

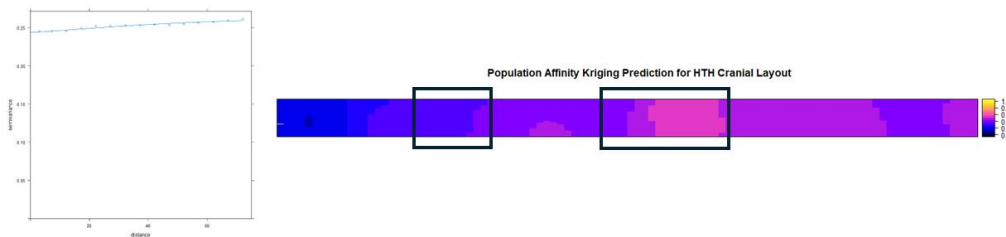


Figure 3.5. Spatial analysis of population within the cranial layout at the HTH Collection. Left: the empirical variogram (dots) and fitted variogram model (line; nugget effect = 0.24; model = stable, psill = 0.02; range = 67.69). Right: kriged map of population, generated using the fitted variogram model. Concentrations of European Americans are dark blue, concentrations of African Americans are purple, and concentrations of Asian Americans and Native Americans are pink. Black rectangles highlight areas of high (left) and medium (right) sampling density as identified in Figure 3.3.

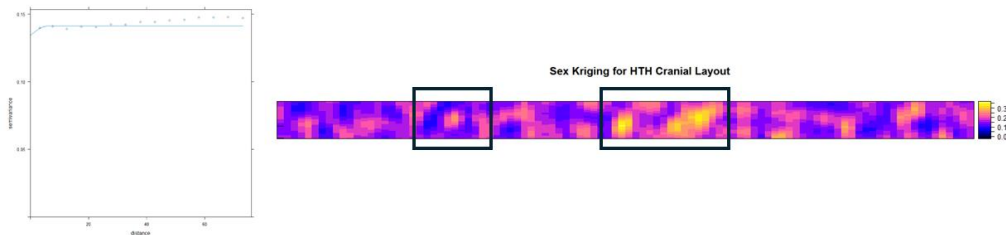


Figure 3.6. Spatial analysis of sex within the cranial layout at the HTH Collection. Left: the empirical variogram (dots) and fitted variogram model (line; nugget effect = 0.13; model = spherical, psill = 0.007; range = 6.02). Right: kriged map of sex, generated using the fitted variogram model. Concentrations of males are dark blue, and concentrations of females are yellow. Black rectangles highlight areas of high (left) and medium (right) sampling density as identified in Figure 3.3.

HTH Collection – Postcranial Layout

Of the ten studies we identified using postcranial samples from the HTH Collection, 1,243 accession numbers were available for analysis. In total, 888 individuals (29.9%) were sampled (Figure 3.7B).

The geospatial analysis of the postcranial layout in relation to sampling frequency at the HTH Collection is presented in Figure 3.7. The experimental variogram model contains a small relative nugget effect (*nugget effect* = 0.26; *psill* = 6.75), accounting for 3.7% of the total variance.

The fitted stable variogram model has a range of 2,738.54 cabinets (Figure 3.7A). Kriging interpolation identified higher sampling density for individuals at the beginning (left) of the collection, as well as those individuals on lower shelves (Figure 3.7C).

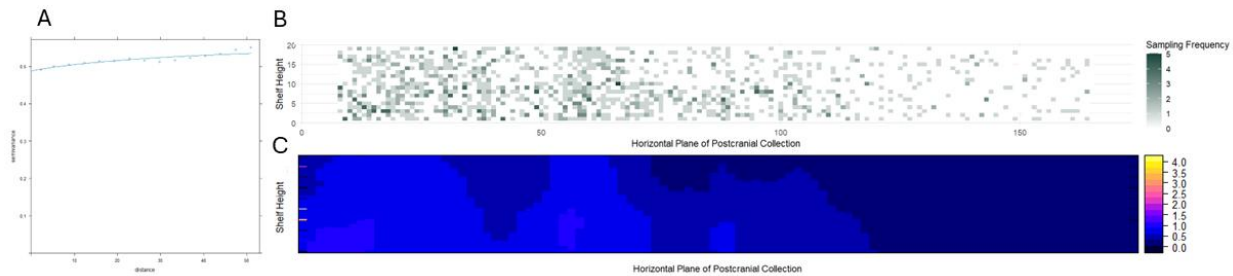


Figure 3.7. *Spatial analysis of postcranial sampling from the HTH Collection. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the HTH postcranial shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by light blue.*

Results of the t-test and one-way ANOVAs are summarized in Table 3.4. Females were sampled at a higher frequency than males; among the nineteen individuals with the highest sample counts ($n = 4$ or $n = 5$), fifteen were female. Sampling frequency was also influenced by shelf height and cabinet number. Individuals located at the beginning of the collection were

significantly more likely to be sampled (see Figure 3.7), as were those stored on lower shelves (Figure 3.8).

Table 3.4. Results of *t*-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency for postcranial samples at the HTH Collection.

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = 19.77	3,678	<0.01	N/A	
Population	ANOVA	F = 23.66	3	<0.01	African American v. European American	<0.01
Population and Sex	ANOVA	F = 76.81	6	<0.01	African American females v. African American males and European American males	<0.01
					European American males v. African American males and European American females	<0.01
					European females v. African American males	<0.01
					0-10 v. 11-19, 20-29, 30-39, 40-49, 70-79, and 80-89	<0.01
Age (Decade)	ANOVA	F = 36.51	10	<0.01	10-19 v. 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89	<0.01
					20-29 v. 30-39, 40-49, 50-59, 60-69, and 70-79	<0.01
					30-39 v. 40-49, 50-59, 60-69, and 70-79	<0.01
					40-49 v. 50-59	<0.05
					40-49 v. 60-69	<0.01
					80-89 v. 50-59 and 70-79	<0.01
					Shelf 3 v. Shelf 13	<0.01
Shelf Height (1-19)	ANOVA	F = 2.77	18	<0.01	Shelf 8 v. Shelf 12, Shelf 13, Shelf 18, and Shelf 19	<0.05
Cabinet Number (1-9)	ANOVA	F = 38.59	8	<0.01	Cabinet 1 v. Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 2 v. Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 3 v. Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 4 v. Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 5 v. Cabinet 7 and Cabinet 8	<0.01
					Cabinet 6 v. Cabinet 8	<0.05

*Only significant pairwise comparisons included for significant ANOVAs.

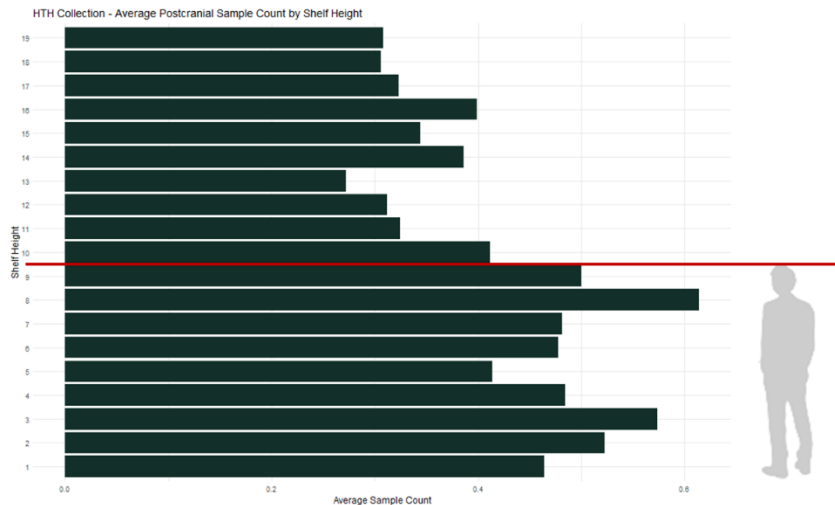


Figure 3.8. Sampling frequencies of postcranial elements by shelf height (lowest shelf: 1; highest shelf: 19). Note: red line indicates approximately 6 feet from ground surface. Adult silhouette is ~ 6'0".

Terry Collection

Of the eight studies we identified using the Terry Collection, 2,194 accession numbers were available for analysis. In total, 1,259 individuals (72.9%) were sampled across these studies (Figure 3.9B).

The geospatial analysis of the Terry Collection physical layout by sampling frequency is presented in Figure 3.9. The experimental variogram model contains a substantial relative nugget effect (*nugget effect* = 0.82; *psill* = 0.52), accounting for 61.2% of the total variance.

The fitted stable variogram model has a range of 31.17 cabinets (Figure 3.9A). Kriging revealed higher sampling densities in the first half of the collection and near the center (Figure 3.9C).

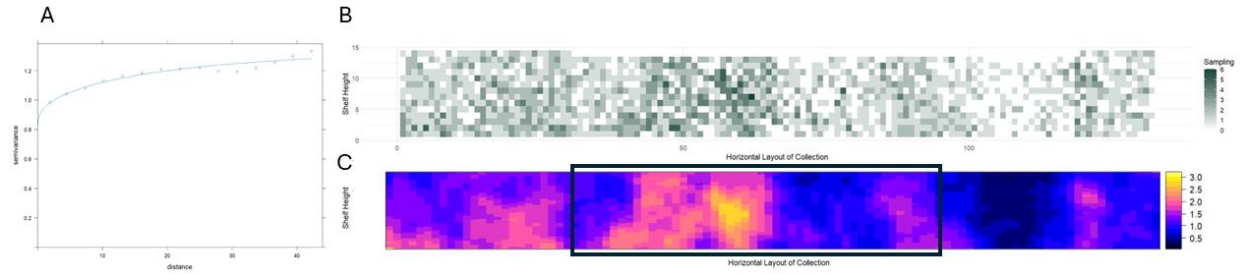


Figure 3.9. Spatial analysis of sampling from the Terry Collection. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the Terry Collection shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by pink and yellow. Rectangle indicates approximate location of cabinets within the analytical room.

Results of the t- tests and one-way ANOVAs are provided in Table 3.5. Males are sampled at a higher frequency than females. Both shelf height and cabinet number influence sampling frequency. Individuals in the first half of the collection are more likely to be sampled (see Figure 3.9), as well as those on lower shelves (Figure 3.10A). Individuals on the shelves within the analytical room were sampled at significantly higher rates ($p < 0.01$) than those in hallways (see Figure 3.10B). Among the ten individuals with the highest sample counts ($n = 5$ or $n = 6$), eight of the ten were female and nine of the ten were located in the analytical room.

Table 3.5. Results of t-tests, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency at the Terry Collection.

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = -5.82	1,059	<0.01	N/A	
Population	ANOVA	F = 16.23	3	<0.01	African American v. European American	<0.01
Population and Sex	ANOVA	F = 17.21	4	<0.01	European American males v. African American males, African American females, and European American females	<0.01
Age (Decade)	ANOVA	F = 11.27	8	<0.01	11-19 v. 60-69 and 70-79	<0.01
					20-29 v. 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89	<0.01
					30-39 v. 60-69 and 70-79	<0.01
					40-49 v. 60-69 and 70-79	<0.01
					50-59 v. 70-79	<0.05
Shelf Height (1-14)	ANOVA	F = 2.29	13	<0.01	Shelf 2 v. Shelf 12 and Shelf 13	<0.05
Cabinet Number (1-9)	ANOVA	F = 37.28	8	<0.01	Cabinet 1 v. Cabinet 4, Cabinet 6, Cabinet 7, and Cabinet 8	<0.01
					Cabinet 2 v. Cabinet 3, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 2 v. Cabinet 4 and Cabinet 5	<0.05
					Cabinet 3 v. Cabinet 4 and Cabinet 8	<0.01
					Cabinet 3 v. Cabinet 7	<0.05
					Cabinet 4 v. Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 7 v. Cabinet 5 and Cabinet 8	<0.05
					Cabinet 8 v. Cabinet 5, Cabinet 6, and Cabinet 9	<0.01
Inside Analytical Room v. In Hallways	t-test	t = -5.85	1,728	<0.01	N/A	

*Only significant pairwise comparisons included for significant ANOVAs.

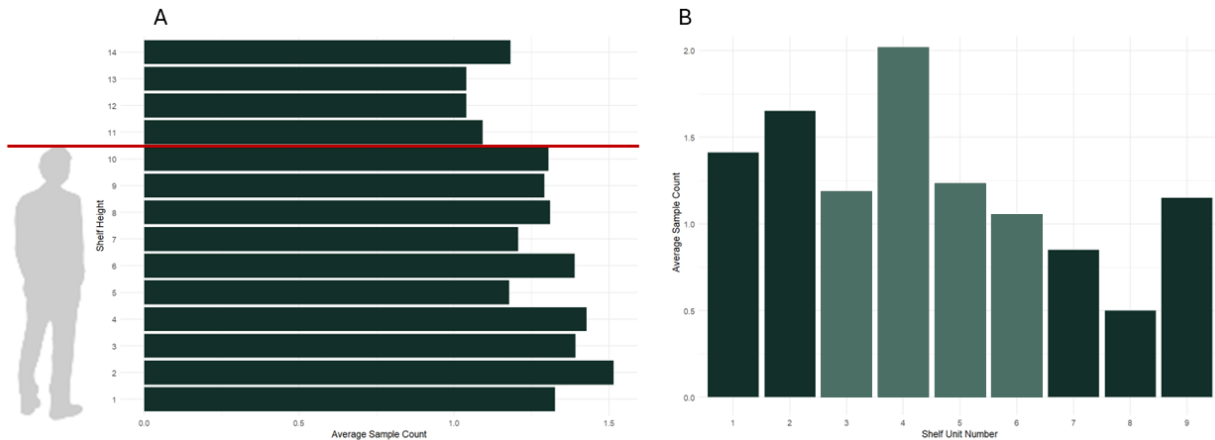


Figure 3.10. (A) Sampling frequencies by shelf height (lowest shelf: 1; highest shelf: 14). (B) Sampling frequencies by cabinet (cabinets within analytical space are light green, while cabinets in hallways are dark green). Note: red line indicates approximately 6 feet from ground surface. Adult silhouette is ~ 6'0".

Areas of high sampling frequencies not explained by location (i.e., within the analytical room versus the hallway) correspond to areas of higher female concentrations, located at either end of the collection (Figure 3.11).

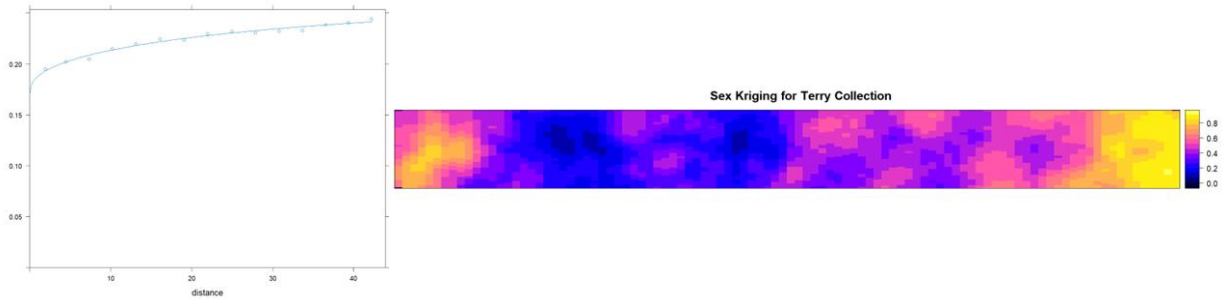


Figure 3.11. Spatial analysis of sex within the Terry Collection. Left: the empirical variogram (dots) and fitted variogram model (line; nugget effect = 0.17; model = stable, psill = 0.094; range = 55.42, kappa = 0.2). Right: kriged map of sex, generated using the fitted variogram model. Concentrations of males are indicated by dark blue, and concentrations of females are indicated by yellow.

UTK DSC – Old Layout (Prior to 2017)

Of the 22 studies we identified using the UTK DSC while in the old layout (prior to 2017), 2,826 accession numbers were available for analysis. In total, 870 individuals (69.0%) were sampled across these studies (Figure 3.12B).

The geospatial analysis of the old UTK DSC physical layout by sampling frequency is presented in Figure 3.12. The experimental variogram model contains a very small relative nugget effect (*nugget effect* = 2.11; *psill* = 678.09), accounting for 0.31% of the total variance.

The fitted stable variogram model has a range of 678.09 cabinets (Figure 3.12A). Kriging identified higher sampling density for individuals in the first half of the collection (Figure 3.12C).

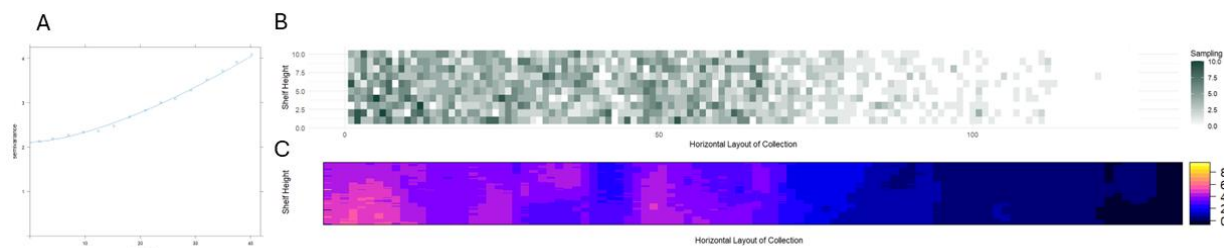


Figure 3.12. Spatial analysis of sampling from the old layout of the UTK DSC. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the UTK DSC shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by pink.

Results of the t-test and one-way ANOVAs are summarized in Table 3.6. African Americans were sampled at higher rates than individuals from other population groups, potentially explaining the higher sampling frequencies observed in the center of the collection (Figure 3.13). Cabinet number was the most influential factor on sampling frequency, as cabinets at the beginning of the collection exhibit significantly higher sampling counts. This pattern remained significant even when accounting for the year of the study (Table 3.7) and is consistent across the various studies, even those conducted in the same year.

Table 3.6. Results of *t*-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency at the old UTK DSC layout (prior to 2017).

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = -1.733	994	0.08	N/A	
Population	ANOVA	F = 5.28	5	<0.01	African American v. European American, Hispanic, and Native American	<0.01
Population and Sex	ANOVA	F = 4.24	9	<0.01	African American males v. Hispanic males, Native American males, European American females, and European American males	<0.01
Age (Decade)	ANOVA	F = 5.54	9	<0.01	30-39 v. 60-69 and 70-79	<0.05
					40-49 v. 60-69 and 70-79	<0.01
Shelf Height (1-10)	ANOVA	F = 0.47	9	0.89	N/A	
Cabinet Number (1-14)	ANOVA	F = 109.8	13	<0.01	Cabinet 1 v. Cabinet 4, Cabinet 5, Cabinet 7, Cabinet 8, Cabinet 9, Cabinet 10, Cabinet 11, Cabinet 12, Cabinet 13, and Cabinet 14	<0.01
					Cabinet 1 v. Cabinet 3	<0.05
					Cabinet 2 v. Cabinet 5, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 3 v. Cabinet 8 and Cabinet 9	<0.01
					Cabinet 4 v. Cabinet 8 and Cabinet 9	<0.01
					Cabinet 5 v. Cabinet 6 and Cabinet 9	<0.01
					Cabinet 6 v. Cabinet 8 and Cabinet 9	<0.01
					Cabinet 7 v. Cabinet 8	<0.05
					Cabinet 9 v. Cabinet 7 and Cabinet 8	<0.01
					Cabinet 10 v. Cabinet 2, Cabinet 3, Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, and Cabinet 8	<0.01
					Cabinet 10 v. Cabinet 14	<0.05
					Cabinet 11 v. Cabinet 2, Cabinet 3, Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 12 v. Cabinet 2, Cabinet 3, Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 13 v. Cabinet 2, Cabinet 3, Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01
					Cabinet 14 v. Cabinet 2, Cabinet 3, Cabinet 4, Cabinet 5, Cabinet 6, Cabinet 7, Cabinet 8, and Cabinet 9	<0.01

*Only significant pairwise comparisons included for significant ANOVAs.

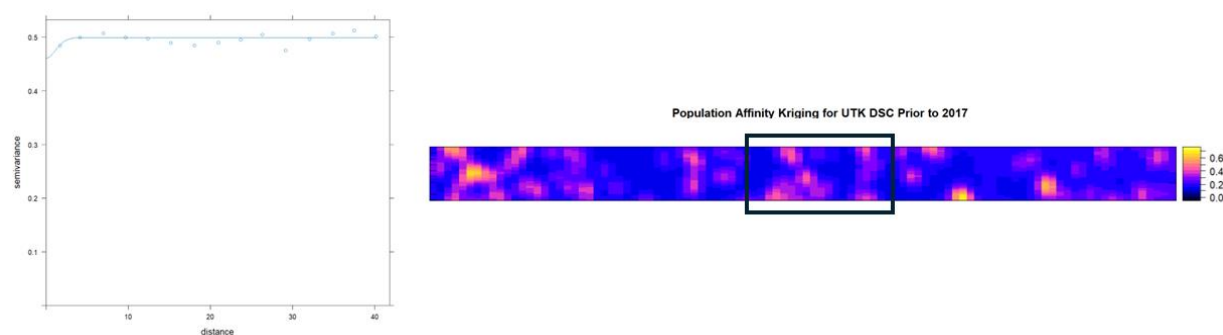


Figure 3.13. Spatial analysis of population affinity within the old UTK DSC layout. Left: the empirical variogram (dots) and fitted variogram model (line; nugget effect = 0.46; model = stable, psill = 0.04; range = 1.73, kappa = 10). Right: kriged map of population affinity, generated using the fitted variogram model. Concentrations of European Americans are indicated by dark blue, and concentrations of African Americans are indicated by pink and yellow. Black square indicates areas of high sampling density in center of collection from Figure 12.

Table 3.7. Results of two-way ANOVA testing the effects of Cabinet Number, Year, and their interaction (Cabinet Number:Year) on sampling counts at the old UTK DSC layout.

Factor of Comparison	DF	Sum Sq	Mean Sq	F Value	Pr (>F)
Cabinet Number	13	245.8	18.91	173.80	<0.01
Year	12	154.4	12.87	118.23	<0.01
Cabinet Number:Year	156	639.7	4.10	37.69	<0.01
Residuals	16,198	1,762.5	0.11	-	-

UTK DSC – Current Layout

Of the seven studies we identified using the UTK DSC in the current layout, 932 accession numbers were available for analysis. In total, 738 individuals (42.8%) were sampled across these studies (Figure 3.14B).

The geospatial analysis of the current UTK DSC collection layout in relation to sampling frequency is presented in Figure 3.14. The experimental variogram model contains a substantial relative nugget effect (*nugget effect* = 0.50; *psill* = 0.01), accounting for 97.6% of the total variance.

The fitted stable variogram model has a range of only 3.49 cabinets (Figure 3.14A). Kriging interpolation identified higher sampling density for individuals in the first three-fourths

of the collection (Figure 3.14C).

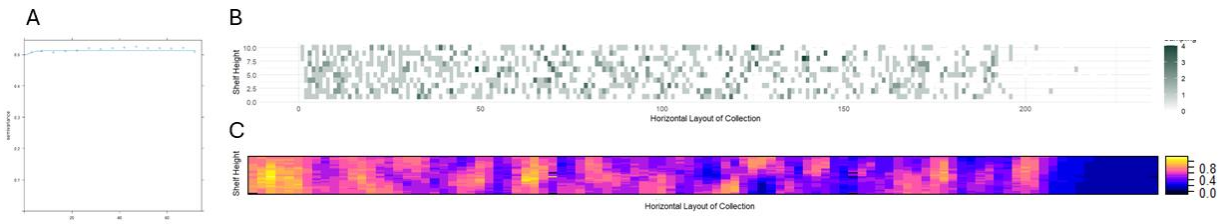


Figure 3.14. Spatial analysis of sampling from the current layout of the UTK DSC. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the UTK DSC shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by pink and yellow.

Results of the t-test and one-way ANOVAs are provided in Table 3.8. In the current layout, females are sampled at a higher frequency than males. Again, African American individuals are sampled at higher rates, explaining some of the areas of high sampling frequencies (Figure 3.15). Cabinet number still influences sampling frequency, even after accounting for the year of the study (Table 3.9).

Table 3.8. Results of t-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency at the current UTK DSC layout.

Factor of Comparison	Test Type	Statistic	DF	P-Value	Tukey Pairwise Comparisons*	Adjusted P-Value
Sex	t-test	t = 5.55	1,172	<0.01	N/A	
Population	ANOVA	F = 18.9	5	<0.01	African American v. European American and Hispanic	<0.01
Population and Sex	ANOVA	F = 15.6	9	<0.01	African American females v. Hispanic males, European American females, and European American males	<0.01
					African American males v. Hispanic males, European American females, and European American males	<0.01
					European American females v. European American males	<0.01
Age (Decade)	ANOVA	F = 1.38	9	0.19	N/A	
Shelf Height (1-10)	ANOVA	F = 1.27	9	0.25	N/A	
Cabinet Number (1-14)	ANOVA	F = 3.68	8	<0.01	Cabinet 1 v. Cabinet 7 and Cabinet 9	<0.01
					Cabinet 1 v. Cabinet 6	<0.05
					Cabinet 2 v. Cabinet 9	<0.05
					Cabinet 3 v. Cabinet 9	<0.05

*Only significant pairwise comparisons included for significant ANOVAs.

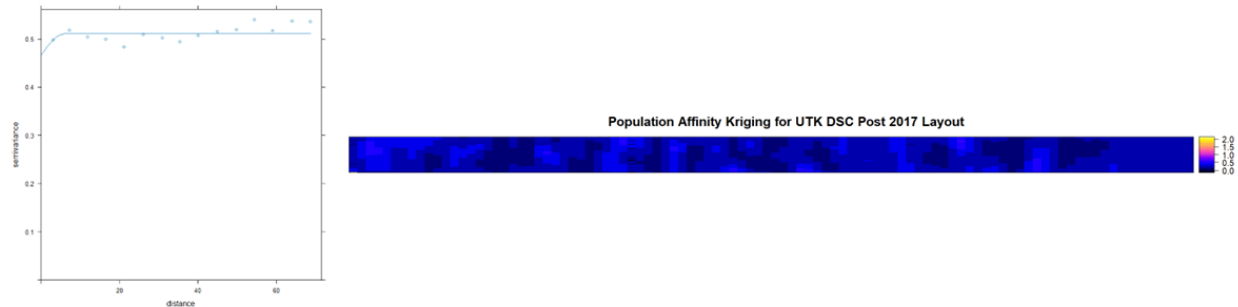


Figure 3.15. Spatial analysis of population affinity within the current UTK DSC layout. Left: the empirical variogram (dots) and fitted variogram model (line; nugget effect = 0.47; model = spherical, psill = 0.04; range = 5.87). Right: kriged map of population affinity, generated using the fitted variogram model. Concentrations of European Americans are indicated by dark blue, and concentrations of African Americans are indicated by light blue.

Table 3.9. Results of two-way ANOVA testing the effects of Cabinet Number, Year, and their interaction (Cabinet Number:Year) on sampling counts at the current UTK DSC layout.

Factor of Comparison	DF	Sum Sq	Mean Sq	F Value	Pr (>F)
Cabinet Number	8	3.7	0.46	4.23	<0.01
Year	3	12.0	4.00	36.83	<0.01
Cabinet Number:Year	24	80.8	3.37	31.00	<0.01
Residuals	6,868	745.7	0.11	-	-

MMDSC

Of the five studies we identified using the MMDSC, 678 accession numbers were available for analysis. In total, 278 individuals (88.8%) were sampled across these studies (Figure 3.16B).

The geospatial analysis of the collection layout in relation to sampling frequency at the MMDSC is presented in Figure 3.16. The experimental variogram model contains a moderate relative nugget effect ($nugget\ effect = 1.18$; $psill = 1.09$), accounting for 52.0% of the total variance.

The fitted stable variogram model has a range of 10.19 cabinets (Figure 3.16A). Kriging interpolation identified higher sampling density for individuals in the first two-thirds of the collection (Figure 3.16C).

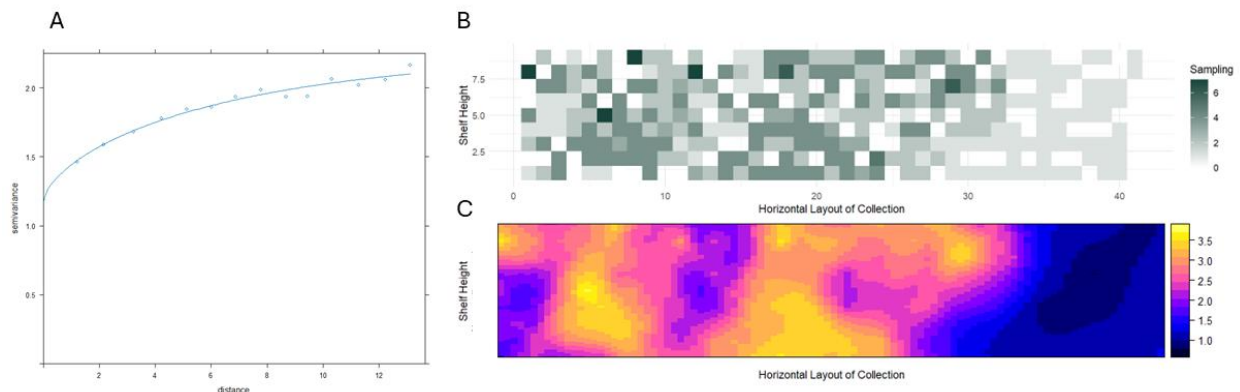


Figure 3.16. Spatial analysis of sampling from the MMDSC. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the MMDSC shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by yellow.

Results of the t-test and one-way ANOVAs are provided in Table 3.10. African American males are sampled at higher rates (a group that includes the four most sampled individuals ($n = 7$)). Cabinet number influences sampling, even after accounting for study year (Table 3.11); individuals to the right of the collection are less likely to be sampled. The areas of low sampling frequencies identified along the bottom left and top middle of the collection map (see Figure 3.16C) are clusters of embalmed individuals (with remnants of adhering tissue) or concentrations of unidentified individuals.

Table 3.10. Results of t-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency at the current UTK DSC layout.

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = -1.97	278	0.05	N/A	
Population	ANOVA	F = 6.15	3	<0.01	African American v. European American	<0.01
Population and Sex	ANOVA	F = 6.19	6	<0.01	African American males v. African American females, Hispanic males, European American females, and European American males	<0.01
					African American males v. Asian males and Hispanic females	<0.05
Age (Decade)	ANOVA	F = 2.35	9	<0.05	50-59 v. 70-79 and 90-99	<0.05
Shelf Height (1-9)	ANOVA	F = 0.97	8	0.46	N/A	
Cabinet Number (1-14)	ANOVA	F = 14.8	6	<0.01	Cabinet 1 v. Cabinet 5 and Cabinet 6	<0.01
					Cabinet 2 v. Cabinet 5 and Cabinet 6	<0.01
					Cabinet 3 v. Cabinet 5, Cabinet 6, and Cabinet 7	<0.05
					Cabinet 3 v. Cabinet 4	<0.05
					Cabinet 4 v. Cabinet 5 and Cabinet 6	<0.01

*Only significant pairwise comparisons included for significant ANOVAs.

Table 3.11. Results of two-way ANOVA testing the effects of Cabinet Number, Year, and their interaction (Cabinet Number:Year) on sampling counts at the MMDSC.

<i>Factor of Comparison</i>	<i>DF</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F Value</i>	<i>Pr (>F)</i>
Cabinet Number	6	16.45	2.74	20.38	<0.01
Year	4	126.81	31.70	235.63	<0.01
Cabinet Number:Year	24	32.94	1.37	10.20	<0.01
Residuals	1,435	193.07	0.13	-	-

TXSTDSC

Of the ten studies we identified using the TXSTDSC, 990 accession numbers were available for analysis. In total, 426 individuals (72.4%) were sampled across these studies (Figure 3.17B).

The geospatial analysis of the collection layout in relation to sampling frequency at the TXSTDSC is presented in Figure 3.17. The experimental variogram model contains a small relative nugget effect ($nugget\ effect = 1.65$; $psill = 23.94$), accounting for 6.4% of the total variance.

The fitted stable variogram model has a range of 140.29 cabinets (Figure 3.17A). Kriging identified higher sampling density for individuals in the first half of the collection (Figure 3.17C).

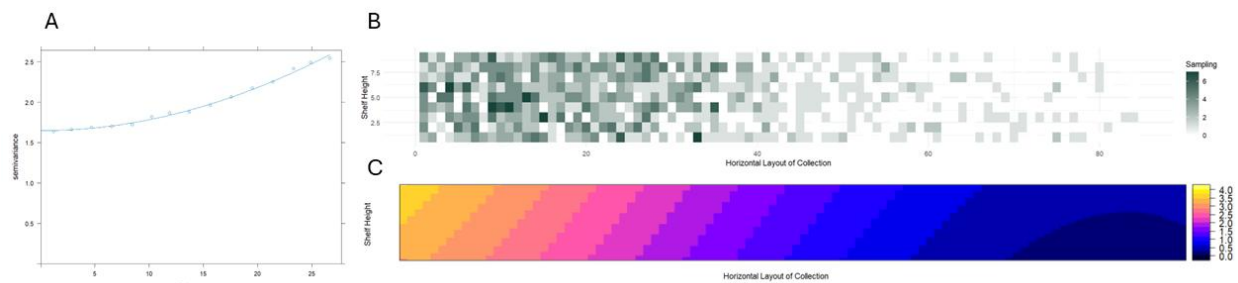


Figure 3.17. Spatial analysis of sampling from the TXSTDSC. (A) The empirical variogram (dots) and fitted variogram model (line) for sampling frequencies. (B) Heatmap illustrating the spatial distribution of sampling frequency across the TXSTDSC shelves. Higher sampling values are represented by dark green. (C) Kriged map of sampling frequency, generated using the fitted variogram model. Lower sampling values are indicated by dark blue and higher sampling values are indicated by yellow.

Results of the t-test and one-way ANOVAs are provided in Table 3.12. Sampling is not influenced by sex or population. The most important influential factor is cabinet number even when accounting for study year (Table 3.13). Individuals at the start of the collection are sampled more frequently than others (see Figure 3.17).

Table 3.12. Results of t-test, one-way ANOVAs, and post hoc Tukey tests for factors influencing sampling frequency at the current UTK DSC layout.

<i>Factor of Comparison</i>	<i>Test Type</i>	<i>Statistic</i>	<i>DF</i>	<i>P-Value</i>	<i>Tukey Pairwise Comparisons*</i>	<i>Adjusted P-Value</i>
Sex	t-test	t = -1.05	476	0.29	N/A	
Population	ANOVA	F = 1.45	10	0.16	N/A	
Population and Sex	ANOVA	F = 1.20	16	0.27	N/A	
Age (Decade)	ANOVA	F = 5.43	9	<0.01	40-49 v. 60-69, 70-79, and 80-89	<0.01
					50-59 v. 70-79 and 80-89	<0.01
					50-59 v. 60-69	<0.05
Shelf Height (1-9)	ANOVA	F = 0.95	8	0.48	N/A	
Cabinet Number (1-3)	ANOVA	F = 195.1	2	<0.01	Cabinet 1 v. Cabinet 2 and Cabinet 3	<0.01

*Only significant pairwise comparisons included for significant ANOVAs.

Table 3.13. Results of two-way ANOVA testing the effects of Cabinet Number, Year, and their interaction (Cabinet Number:Year) on sampling counts at the TXSTDSC.

<i>Factor of Comparison</i>	<i>DF</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F Value</i>	<i>Pr (>F)</i>
Cabinet Number	2	154.3	77.17	208.5	<0.01
Year	3	395.1	131.69	355.9	<0.01
Cabinet Number:Year	6	419.9	69.99	189.1	<0.01
Residuals	2,340	865.9	0.37	-	-

DISCUSSION

This paper provides valuable insights into how the physical layout of a documented skeletal collection may introduce selection bias in biological anthropological research. Our findings demonstrate spatial layout plays a significant role in shaping sampling practices.

Across all collections, cabinet number consistently influenced sampling; individuals at the beginning of the collection are always sampled at higher frequencies. While individuals at the beginning of a collection would be sampled more frequently because they have been available for analysis longer, this pattern persisted in the historic collections that ceased accepting individuals before the studies examined were conducted. Similarly, in modern collections, this trend remained significant even when accounting for the potentially confounding effects of study year. So, some other factor is causing this phenomenon. Perhaps time does play a

role. Consider Bethard (2005:20), who evaluated skeletons in order of accession number “until time constraints prohibited additional data collection.” This suggests skewed sampling distributions may reflect researchers starting with the earlier individuals and running out of time before reaching the end of the collection. If true, this spatial trend may be better explained by time management than by collection layout alone.

Nevertheless, the physical layout of the collection does play a role in shaping sampling patterns. Consider, for example, shelf height and cabinet location. In the Terry Collection and the postcranial layout of the HTH Collection, the collections with the tallest cabinets, individuals stored on higher shelves were sampled significantly less frequently than those on lower shelves. The demographic data cannot explain this pattern, so some individuals are under sampled purely due to a less accessible position within the skeletal collection. Similarly, the unique layout of the Terry Collection highlights the influence collection layout can have. Cabinets in the analytical room were sampled at significantly higher rates than those located in the hallway. When higher sampling frequencies outside of the analysis room are detected, they are usually associated with higher concentrations of the females located at either end of the Terry Collection. This finding is due in large part to Mildred Trotter, who worked to balance the sex ratio in the collection (Hunt and Albanese 2005; Wilson 2022). In other words, researchers are more likely to sample from more convenient locations unless specific demographic parameters are needed to meet study demands.

Modern collections appear less influenced by these physical layout-related sampling biases. Cabinets in these collections are all located in the same room and their shelves are shorter than the historic collections. Although the distance from collection cabinets to analytical space could still play a role—since longer distances may increase the time required to evaluate each

individual—this study did not collect information on those logistical factors (e.g., target sample size versus number of individuals actually evaluated). The push for more modern documented skeletal collections in research efforts is directly related to the more representative nature of these collections to contemporary populations (Albanese 2018), so improved convenience may reduce selection bias in future research. However, researchers will always contend with inherent collection biases, such as an underrepresentation of females. These biases persist regardless of sampling strategy, forcing researchers to address these additional sources of error, including selection bias. There is also the need to ensure the same few individuals in the early accession numbers of a collection are not repeatedly used when other suitable options are available. This practice has the potential to bias the representation of human variation in research to a few individuals.

The primary limitation of this study is the small sample size due to limited survey responses. Additional sampling data from multiple studies would provide a more robust understanding of these trends. However, while preliminary, this study demonstrates the potential of geostatistical and spatial analysis as tools to explore collection layout and how physical space could inadvertently introduce selection bias. While the figures in this study should not be used to make concrete decisions regarding physical collection, sampling strategies, etc., they do visually assess why certain individuals are sampled more frequently than others. To the best of our knowledge, geostatistics and spatial analysis has not yet been applied in this context.

These results highlight the importance of considering spatial organization when designing research protocols and interpreting findings from skeletal collections. Sampling strategies should consider the unique problem of data collection from skeletal collections, and bear in mind some of the trappings that have previously led to under and over sampling certain individuals.

Although some collections provide randomized sample lists (for example, UTK DSC offers custom random samples based on study criteria), this is a service generally at the researcher's request. Curators and collection managers cannot anticipate the specific sampling needs of every research study, placing the onus on researchers to implement proper strategies, including requesting random samples that meet their study criteria. And, to avoid oversampling individuals located at the beginning of the collection, researchers can divide randomized samples into subsets and process each list sequentially. This approach ensures a broader distribution of sampled individuals across the collection. Researchers should also make a concerted effort to include individuals located on higher shelves, rather than omitting them due to accessibility challenges. Addressing these biases will improve the representativeness of sampled individuals, ensure equitable utilization of skeletal collections, and enhance the validity of research outcomes.

CONCLUSION

This study underscores the significant role spatial layouts play in shaping sampling practices in research using documented skeletal collections. Cabinet location and shelf accessibility, such as height and proximity to analytical spaces, were identified as key factors influencing sampling frequencies. In addition, this study revealed that individuals at the beginning of collections are sampled at significantly higher rates. While these spatial biases can partly be attributed to the physical layouts of each collection, researcher practices—such as evaluating individuals in chronological order until time runs out—likely contribute to the uneven sampling distribution. Researchers need to employ sampling strategies to specifically mitigate effects of collection layout and must sample all individuals, regardless of their location within collection cabinets. To further explore selection bias at documented collections, future studies

should conduct similar analyses with larger samples and more demographic information for individuals within each collection.

BIBLIOGRAPHY

- Albanese, J. (2018). Strategies for dealing with bias in identified reference collections and implications for research in the 21st century. In: C. Y. Henderson & F. A. Cardoso, eds. *Identified skeletal collections: The testing ground of anthropology?* Oxford, UK: Archaeopress Publishing.
- Borcard, D., Gillet, F., & Legendre, P. 2018. *Numerical ecology with R, second edition*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-319-71404-2>
- Body Donation. 2024. Forensic Anthropology Center, University of Tennessee, Knoxville. Retrieved October 13, 2024, from <https://fac.utk.edu/body-donation/>
- Campanacho, V., Ales Cardoso, F., & Ubelaker, D. H. 2021. Documented skeletal collections and their importance in forensic anthropology in the United States. *Forensic Sciences* 1:228-239. <https://doi.org/10.3390/forensicsci1030021>
- Collections, 2024. Cleveland Museum of Natural History. Retrieved November 22, 2024, from <https://www.cmnh.org/science-conservation/areas-of-study/anthropological-sciences/collections>
- Forensic Anthropology Center. 2022. Guidelines for Collection Research. Retrieved November 22, 2024, from <https://fac.utk.edu/wp-content/uploads/2022/12/Guidelines-for-Collections-2022.pdf>
- Gocha, T. P., Mavroudas, S. R., & Wescott, D. J. 2022. The Texas State Donated Skeletal Collection at the Forensic Anthropology Center at Texas State. *Forensic Sciences* 2(1):7-19. <https://doi.org/10.3390/forensicsci2010002>
- Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W., & Heuvelink, G. B. M. 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. *Computers & Geosciences*, 35(8): 1711-1721. <https://doi.org/10.1016/j.cageo.2008.10.011>
- Hefner, J. T. 2017. Biological distance and geospatial analysis. In: M. Heilen, J. T. Hefner, & M. A. Keur, eds. *Deathways and lifeways in the American Southwest: Tucson's historic Alameda-Stone Cemetery and the transformation of a remote outpost into an urban city, Volume 2*. Technical Report 10-96: Statistical Research Inc. pp 399-458.
- Hunt, D. R., & Albanese, J. 2005. History and demographic composition of the Robert J. Terry Collection. *American Journal of Physical Anthropology* 127:406-417.
- Komar, D., & Grivas, D. 2008. Manufactured populations. What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum Documented Collection. *American Journal of Biological Anthropology* 137(2): 244-233. <https://doi.org/10.1002/ajpa.20858>

- Legendre, P., & Legendre, L. 1998. *Numerical ecology, second edition*. Amsterdam, The Netherlands: Elsevier Science B.V.
- Muller, J. L., Pearlstein, K. E., & de la Cova, C. 2017. Dissection and documented skeletal collections: Embodiments of legalized inequality. In: K. C. Nystrom, ed. *The bioarchaeology of dissection and autopsy in the United States*. New Paltz, NY: Springer. pp. 185-201.
- Oliver, M. A., & Webster, R. 2015. *Basic steps in geostatistics: The variogram and kriging*. Reading, UK: Springer.
- Pebesma, E. J. 2004. Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30(7): 683-691. <https://doi.org/10.1016/j.cageo.2004.03.012>
- Pebesma, E. J., & Bivand, R. S. 2005. Classes and methods for spatial data in R. *R News*, 5(2): 9-13. <https://journal.r-project.org/articles/RN-2005-014/RN-2005-014.pdf>
- R Core Team. 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Relethford, J. H. 2008. Geostatistics and spatial analysis in biological anthropology. *American Journal of Physical Anthropology* 136:1-10.
- Sholts, S. B. 2024. "To honor and remember": An ethical awakening to African American remains in museums. *American Journal of Biological Anthropology*, Early View:e24943. <https://doi.10.1002/ajpa.24943>
- Smith, J. R. 2019. Living with observational data in biological anthropology. *American Journal of Biological Anthropology* 169: 591-598.
- Smith, J., & Noble, H. 2014. Bias in research. *Evidence Based Nursing* 17(4): 100-101.
- Wilson, E. 2022. *Mildred Trotter and the Invisible Histories of Physical and Forensic Anthropology*. Boca Raton, FL: CRC Press.

APPENDIX A: ANONYMOUS SURVEY QUESTIONS

Sampling and Selection Bias in Research Using Documented Skeletal Collections

Research Participant Information and Consent Form

Investigators: Rhian Dunn¹, M.S.; Joseph T. Hefner¹, Ph.D.

¹Department of Anthropology, Michigan State University

As a valuable member of the forensic anthropology community, you are invited to participate in a research study investigating sampling and selection bias in forensic anthropological research using United States-based documented skeletal collections. The purpose of this research is to investigate several potential factors that introduce selection bias. The results of this study will assess the extent of bias, if present, in an effort to understand the true randomization of study samples.

This study will include participants in the field of forensic anthropology, who are 18 years of age or older, and who have conducted research in a United States-based documented skeletal collection. To participate in this study, you will be asked to complete an anonymous online survey. As this survey is anonymous, no records identifying study participants will be maintained. Your participation is voluntary and you can skip any question you do not wish to answer or exit from the survey at any time. Your responses will not be recorded until the survey is submitted. The anticipated time to complete this survey is approximately 10-20 minutes.

This study was reviewed by Michigan State University's Institutional Review Board and was

determined to be exempt. If you have questions or concerns about your role and rights as a research participant, please contact Michigan State University's Human Research Protection Program by phone (517-355-2180) or email (irb@ora.msu.edu).

If you have any concerns or questions about this study, please contact investigators Rhian Dunn (dunnrhia@msu.edu) or Joseph T. Hefner (hefnerj1@msu.edu). This project was supported by Award No. 15PNIJ-23-GG-01939-RESS, awarded by the National Institute of Justice, Office of Justice Programs, U.S Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this study are those of the authors and do not necessarily reflect those of the Department of Justice.

By clicking the "I Consent" option below, you indicate that you voluntarily agree to participate in this research study.

What is your highest degree earned?

- High School Diploma/GED
- BA/BS
- MA/MS
- PhD or equivalent
- Other

How many studies have you participated in that used samples from documented skeletal collections?

- 1-5
- 6-10
- 11-15

- 15+

The following section asks for some basic information about studies you have conducted using **United States-based documented skeletal collections**. These studies do not need to include published works—they can include ongoing studies or unpublished master’s theses and/or doctoral dissertations. This study will not ask for any identifying information.

As part of this survey, we are asking that you share the catalogue numbers of individuals used in your study. **We are not asking for you to share any data, results, analysis, or any other information that belongs to you.** We solely wish to know which individuals you sampled from the collection(s) for your study. Catalogue numbers will NOT be published in any form—we are collecting them for analytical purposes only.

For this survey, you will add information for one study at a time. You can add information on up to 10 studies.

STUDY 1

1. What collection(s) did you use in this research project?
2. In what year did you collect this data? (An approximate year is fine)
3. What was the general research focus in this study? (Age, sex, trauma, etc.)
4. Which elements did you use?
 - a. Skull
 - b. Postcrania
 - c. Both

5. What sampling strategy did you use for this study (e.g., none, matched another dataset, simple random, stratified random, etc.)?
6. Did your study require specific demographic samples (e.g., age cohorts, even sex distribution, particular groups)? If so, what was your study's required criteria?
7. Please provide a list of the catalogue numbers (e.g., HT0001) of all individuals used in your sample for this study. Enter in the text box below or use the file upload (next question). We are looking for the catalogue numbers specific to the collections—please do not submit dummy numbers created for your specific study.

If you used Excel for your data collection, you can copy and paste your ID column into this text box. A keyboard shortcut to select data from an entire Excel column is ctrl+shift+down arrow (cmd+shift+down arrow for macs).

We are not asking for you to share any data, results, analysis, or any other information that belongs to you. We solely wish to know which individuals you sampled from the collection(s) for your study.

8. If you prefer, you can upload a file with your catalog numbers here. We are looking for the catalogue numbers specific to the collections—please do not submit dummy numbers created for your specific study.

Please ONLY share a file that includes your catalog numbers. DO NOT upload files with any other information from your study.

We do not want you to share any data, results, analysis, or any other information that belongs to you. We solely wish to know which individuals you sampled from the collection(s) for your study.

9. Any comments, concerns, or additional information to provide?

We do not want you to share any data, results, analysis, or any other information that belongs to you. Please DO NOT share any identifying information.

10. Do you wish to provide information from another study?

a. Yes

b. No

Study Questions Repeat for up to ten studies.

Do you have any comments you wish to share with the study investigators?

APPENDIX B: INTERVIEW QUESTIONS

Interview Questions for Documented Skeletal Collection Managers or Familiar

Practitioners

1. Who gets access to the donated collection?
2. Is there a minimum or maximum number of individuals that can be used in a sample?
3. Are randomized samples provided to visiting scholars? If so, how are they randomized?
4. How much does the curator(s)/collection manager(s) interact with visiting scholars, or assist in their data selection?
5. Has the collection changed locations since its inception?
6. How long has the donated collection been historically situated in the current layout? (e.g., shelf layout and individual storage order)
 - a. If it has changed, what was/were the layout(s) previously?
7. How were/are new donations added to the collection?
 - a. Is each new donation assigned the next available catalogue number?
 - b. Are catalogue numbers ever re-assigned if an individual is de-accessioned?

CONCLUSION

The purpose of this dissertation was to investigate the presence and extent of bias in research using documented skeletal collections. Bias is an inherent challenge in any research discipline; forensic anthropology is no exception. Previous research has addressed the effect of contextual bias on forensic anthropological casework, but few studies have assessed the impact of biases arising during the initial stages of research when samples are sourced from documented skeletal collections. Given forensic anthropology's reliance on these collections for research samples, understanding bias therein is critical.

This dissertation focused on three specific aspects of research bias: 1) collection-specific sample bias, investigated through a comparison of craniometric data for individuals from eight documented skeletal collections; 2) representation of documented skeletal collections to forensic anthropological cases, assessed using individuals from documented collections and from forensic cases found in the computer program FORDISC; and, 3) selection bias, examined using geostatistical and spatial analyses of data from five documented skeletal collections. These biases were addressed in three separate manuscripts.

Manuscript One—"Collection-Specific Sample Bias in United States-based Documented Skeletal Collections"—investigated the presence of collection-specific sample bias and whether that bias impacts our understanding of human skeletal variation. To address this question, craniometric data from individuals across eight modern U.S.-based documented skeletal collections, including those in geographically similar and distinct regions, were used. Exploratory analyses and multivariate statistical methods were employed to compare individuals across the documented collections, controlling for known demographic variables, such as population and sex. The results revealed statistically significant differences in cranial

morphology between collections, suggesting collection-specific sample bias. However, this bias has a smaller influence on the aggregation of cranial morphology represented at each collection than demographic variables such as sex and population. In sum, this study demonstrated how documented skeletal collections exhibit collection-specific sample bias but also suggests they remain valid and valuable sources of samples for research concerning human skeletal variation and forensic anthropological methods.

Manuscript Two—“*Representation of Documented Skeletal Collections to Forensic Casework*”—investigated the representativeness of documented skeletal collections to forensic anthropological casework. To address this question, craniometric data were collected from individuals across eight modern U.S.-based documented skeletal collections. Additionally, craniometric data from individuals in the Forensic Anthropology Data Bank (FDB) were used as a proxy for U.S. wide forensic cranial diversity. Custom databases were created in Fordisc 3.1 (FD3) to evaluate the inclusion of larger samples of documented skeletal collections and their influence on FD3’s classification accuracy. The results of this study indicate documented skeletal collections are most representative of the populations that are well reflected within their collections. However, the limited demographic diversity in these collections restricts broader applicability for forensic anthropological casework. Furthermore, samples from the University of Tennessee, Knoxville Donated Skeletal Collection have a significant impact on model performance in FD3. These results highlight the need for increased case submissions to the FDB to increase the representativeness of FD3 reference samples and to improve classification performance in forensic casework.

Manuscript Three—“*Geostatistical and Spatial Analysis of Selection Bias in Documented Skeletal Collections*”—investigated selection bias in studies utilizing documented skeletal

collections. To address this question, three types of data were collected: 1) geospatial layouts of five U.S.-based documented skeletal collections; 2) study sample catalogue numbers obtained through an anonymous survey and extensive literature review; and, 3) contextual information on research sampling and collection layouts, gathered from anonymous interviews with individuals familiar with at least one of the five collections. Geostatistical and spatial analyses were used to assess sampling frequency patterns at each collection, while contextual information was used to interpret those patterns. These findings indicate collection layout significantly influences sampling at documented collections; individuals near the beginning of the collection are consistently oversampled and the individuals located on higher shelves are frequently under sampled. These results underscore the need for more rigorous sampling strategies in forensic anthropology to ensure all individuals within a collection have an equal likelihood of inclusion in any one study.

The three manuscripts in this dissertation investigated three potential confounding factors encountered when sampling from U.S.-based documented skeletal collections and identified the presence of collection-specific sample bias, limits in the representativeness of documented skeletal collections to forensic cases, and the existence of selection bias in research sampling. These findings indicate that validation efforts must extend beyond the use of a single documented skeletal collection, as these samples differ significantly from each other; instead, validation studies must incorporate multiple collections to ensure analytical methods are not impacted by sample bias. To increase the representativeness of documented skeletal collections to forensic cases, we must increase the diversity of collection samples; conversely, to reduce sample bias within FD3, we must continue to submit identified forensic cases to the FDB to increase sample sources contributing to FD3 reference samples. Lastly, we must develop and

validate documented skeletal collection-specific sampling strategies to mitigate selection bias and reduce the oversampling of individuals located at the beginning of each collection. One potential strategy is to request a randomized sample list from collection managers and curators that meets study requirements; this list can then be divided into manageable subsets (e.g., based on daily or weekly sampling capacity). The use of subset samples enables researchers to organize each subset chronologically to streamline sampling while still ensuring more broad coverage across the collection within the research visit. This is the sampling approach utilized for this dissertation to ensure random samples, but it has yet to be validated for mitigating selection bias.

Bias research is essential to ensure the reliability and validity of analytical methods used in forensic anthropology and is necessary to meet the *Daubert* guidelines for expert witness testimony. The manuscripts included in this dissertation demonstrate sample and selection biases exist within the research utilizing documented skeletal collections, but these collections remain invaluable resources for research in forensic anthropology. Additionally, these manuscripts provide a foundation for future investigations into biases for the many documented skeletal collections not included in this dissertation, such as those housed internationally. Addressing the bias encountered in research using documented skeletal collections is necessary and achievable, but will likely involve continued evaluation, the development of strategies for sampling from collections, and efforts to increase diversity and representation within documented collections. Ultimately, this dissertation contributes to the advancement of forensic anthropology and hopes to aid in its ability to serve diverse communities with scientific rigor.