biology

Recent Advances in Forensic Anthropological Methods and Research

Edited by
Ann H. Ross and Eugénia Cunha

Printed Edition of the Special Issue Published in *Biology*

www.mdpi.com/journal/biology

MDPI

Recent Advances in Forensic Anthropological Methods and Research

Recent Advances in Forensic Anthropological Methods and Research

Editors

Ann H. Ross
Eugénia Cunha

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Ann H. Ross
Biological Sciences
North Carolina State
University
Raleigh
United States

Eugénia Cunha
Laboratory of Forensic
Anthropology
Centre for Functional Ecology
Department of Life Sciences
University of Coimbra
Coimbra
Portugal

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Biology* (ISSN 2079-7737) (available at: www.mdpi.com/journal/biology/special_issues/forensic_anthropological_methods_research).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-5432-7 (Hbk)
ISBN 978-3-0365-5431-0 (PDF)

Cover image courtesy of Eugénia Cunha

Contents

About the Editors

Ann H. Ross

Ann H. Ross, Ph.D., D-ABFA, is a Board-Certified forensic anthropologist, in the Department of Biological Sciences at NC State. She runs the NC Human Identification and Forensic Analysis Lab at State. Professor Ross has been committed to human rights and disaster work and has been deployed to Bosnia after the genocide, worked for the Panamanian and Chilean Truth Commissions. Dr. Ross is also a member of the Federal DHHS Disaster Mortuary Operational Response Team and assisted in the identifications after Hurricane Katrina and US Citizens who perished in the Haiti Earthquake. Her laboratory also has a large forensic anthropology caseload and has been very active in working with multidisciplinary teams in North Carolina to reduce the backlog of unidentified remains. Professor Ross teaches courses in human identification and cold case investigations, to name a few.

Eugénia Cunha

Eugénia Cunha, Ph.D., C- FASE, is a forensic anthropologist and the Director of the South Delegation of the National Institute of Legal Medicine and Forensic Sciences, Lisbon, Portugal. She is also a full professor at the University of Coimbra since 2003, where she created and co-coordinates the Laboratory of Forensic Anthropology. She is a Co-founder and former President of FASE-Forensic Anthropology Society of Europe (2009–2016); Vice-President and Founder member of ABRAF - Associação Brasileira de Antropologia Forensel; Fellow of the American Academy of Forensic Sciences; Member of the Pathology and Anthropology Sub-group at the Interpol DVI Working Group; Roster member of JRR, Justice Rapid Response. Invited teacher, among others, at Stanford University, USA (Tinker visiting research, Center of Latin American Studies, 2020). Since 1997, she has conducted more than 500 forensic anthropology cases in Portugal and abroad (Brazil and some African countries). She has been a consultant and evaluator to several entities in around 15 different countries and invited speaker in about 25 countries. She is a top peer-reviewer and a member of the Editorial Board of some scientific journals. She is the co-editor/author of four books and author of more than 150 peer-reviewed publications. To date, 22 Ph.D. students have already accomplished their Ph.D. under her supervision. Her research aims are focused on forensic anthropology, specifically in identification.

biology

MDPI

Editorial

Recent Advances in Forensic Anthropological Methods and Research

Eugénia Cunha [1,2] and Ann H. Ross [3,*]

[1] Centre for Functional Ecology (CEF), Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; cunhae@ci.uc.pt
[2] National Institute of Legal Medicine and Forensic Sciences, 1169-201 Lisbon, Portugal
[3] Human Identification & Forensic Analysis Laboratory, Department of Biological Sciences, North Carolina State University, Raleigh, NC 276995, USA
* Correspondence: ahross@ncsu.edu

This Special Issue, "Recent Advances in Forensic Anthropological Methods and Research", with thirteen articles covers a wide range of highly diverse topics within forensic anthropology. Topics ranging from innovative approaches to critical reviews have received much attention, with more than thirteen thousand views during the past year. This is unequivocal proof of the interest in this Special Issue. Authors representing Europe, the United States, Australia, and South Africa embody the breadth of the present-day research being conducted in forensic anthropology.

In regard to estimating biological profiles (e.g., biological sex, age at death, population affinity, and stature), there are three articles focusing on age at death. One manuscript by Niel, Chaumoître, and Adalian [1] addresses bias due to altered growth trajectories in estimating juvenile aging in fetuses and infants. Two manuscripts discuss aging adults, considered to be the Achilles heel of forensic anthropology. A paper by Dias, Manco, Corte Real, and Cunha [2] proposes a blood–bone–tooth model using DNA methylation to predict age in forensic contexts. This paper presents an interesting alternative for aging the dead and the living, and brings new insights into the development of multitissue age prediction models as applied to blood, bone, and teeth. The third adult age estimation article by Navega, Costa, and Cunha [3] proposes a new method based on a multifactorial macroscopic analysis and deep random neural network models. Within the generic factors of identity (i.e., biological profile), the ever-polemic topic of population affinity is discussed and illustrated using geometric morphometric and spatial analysis methods within Latin America. Ross and Williams [4] argue that there is a benefit to and necessity of embracing studies that employ population structure models to better understand human variation and the historical factors that have influenced it.

Within the realm of individualizing factors, Butaric, Richman, and Garvin [5] discuss the potential factors that might affect the reliability of using frontal sinuses for personal identification. Their study investigates how slight deviations in orientations affect sinus size and outline shape, which could potentially impact identification.

New approaches are illustrated by the article by Procopio, Mein, Starace, Bonicelli, and Williams [6], which shows that bone proteomics is a well-founded resource with which to identify microbially driven versus extrinsically driven bone diagenesis. Another novel subject is the review by Marquez-Grant and colleagues on the effects of various drugs on the skeleton, including prescription and recreational drugs, that could affect forensic anthropological analyses [7]. Another new approach by McWhirter and colleagues describes how to accurately individualize skeletons from commingled remains using mesh-to-mesh value comparisons for pair matching skeletal elements [8].

A topic with increasing attention is forensic facial comparison, which is the subject of one paper by Bacci and coworkers that discusses relevant terminology, the validity as well as reliability of the Facial Identification Scientific Working Group's list of morphological features, and proposes standards for CCTV equipment [9].

Citation: Cunha, E.; Ross, A.H. Recent Advances in Forensic Anthropological Methods and Research. *Biology* **2022**, *11*, 908. https://doi.org/10.3390/biology11060908

Received: 9 June 2022
Accepted: 9 June 2022
Published: 13 June 2022

The need to know the attributions of each area of expertise in forensic anthropology is discussed by Passalacqua, Pilloud, and Congram [10], who call attention to ethical procedures and requisite qualifications. Furthermore, they emphasize the need to develop standards and best practice guidelines.

One of the main reasons why forensic anthropologists are called to testify in court is because of traumatic injuries to skeletal tissues. The article of de Boer, Berger, and Blau [11] discusses and examines the concept of 'degree of force' as well as why it is considered a pertinent issue in legal proceedings.

One of the big challenges in skeletal traumatic injuries interpretation is to perform discrimination among BFT and thermal-induced trauma. Keys and Ross [12] conducted an experiment that found that blunt force trauma signatures remained after burning. It concludes that there are distinct patterns attributed to thermal fractures and blunt force fractures.

Nonhuman skeletal remains continue to be part of the routine cases of forensic anthropologists. The Garvin team [13] assesses the utility of quantitative methods for distinguishing human from nonhuman remains and presents additional resources for species identification.

We can consider that we have accomplished our aims of presenting a wide array of methods and topics that are unquestionably relevant to the practice of forensic anthropology. The quality of expertise has to derive from modern and updated research.

A teoria orienta, a experiência decide.

Author Contributions: Conceptualization, E.C. and A.H.R.; writing—original draft preparation, E.C. and A.H.R.; writing—review and editing, E.C. and A.H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Niel, M.; Chaumoître, K.; Adalian, P. Age-at-Death Estimation of Fetuses and Infants in Forensic Anthropology: A New "Coupling" Method to Detect Biases Due to Altered Growth Trajectories. *Biology* **2022**, *11*, 200. [CrossRef] [PubMed]
2. Correia Dias, H.; Manco, L.; Corte Real, F.; Cunha, E. A Blood–Bone–Tooth Model for Age Prediction in Forensic Contexts. *Biology* **2021**, *10*, 1312. [CrossRef] [PubMed]
3. Navega, D.; Costa, E.; Cunha, E. Adult Skeletal Age-at-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis. *Biology* **2022**, *11*, 532. [CrossRef] [PubMed]
4. Ross, A.H.; Williams, S.E. Ancestry Studies in Forensic Anthropology: Back on the Frontier of Racism. *Biology* **2021**, *10*, 602. [CrossRef] [PubMed]
5. Butaric, L.N.; Richman, A.; Garvin, H.M. The Effects of Cranial Orientation on Forensic Frontal Sinus Identification as Assessed by Outline Analyses. *Biology* **2022**, *11*, 62. [CrossRef] [PubMed]
6. Procopio, N.; Mein, C.A.; Starace, S.; Bonicelli, A.; Williams, A. Bone Diagenesis in Short Timescales: Insights from an Exploratory Proteomic Analysis. *Biology* **2021**, *10*, 460. [CrossRef] [PubMed]
7. Márquez-Grant, N.; Baldini, E.; Jeynes, V.; Biehler-Gomez, L.; Aoukhiyad, L.; Passalacqua, N.V.; Giordano, G.; Di Candia, D.; Cattaneo, C. How Do Drugs Affect the Skeleton? Implications for Forensic Anthropology. *Biology* **2022**, *11*, 524. [CrossRef] [PubMed]
8. McWhirter, Z.; Karell, M.A.; Er, A.; Bozdag, M.; Ekizoglu, O.; Kranioti, E.F. Exploring the Functionality of Mesh-to-Mesh Value Comparison in Pair-Matching and Its Application to Fragmentary Remains. *Biology* **2021**, *10*, 1303. [CrossRef]
9. Bacci, N.; Davimes, J.G.; Steyn, M.; Briers, N. Forensic Facial Comparison: Current Status, Limitations, and Future Directions. *Biology* **2021**, *10*, 1269. [CrossRef]
10. Passalacqua, N.V.; Pilloud, M.A.; Congram, D. Forensic Anthropology as a Discipline. *Biology* **2021**, *10*, 691. [CrossRef] [PubMed]
11. De Boer, H.H.; Berger, C.E.H.; Blau, S. Providing a Forensic Expert Opinion on the "Degree of Force": Evidentiary Considerations. *Biology* **2021**, *10*, 1336. [CrossRef] [PubMed]
12. Keys, K.; Ross, A.H. Identifying Blunt Force Traumatic Injury on Thermally Altered Remains: A Pilot Study Using *Sus scrofa*. *Biology* **2022**, *11*, 87. [CrossRef] [PubMed]
13. Garvin, H.M.; Dunn, R.; Sholts, S.B.; Litten, M.S.; Mohamed, M.; Kuttickat, N.; Skantz, N. Forensic Tools for Species Identification of Skeletal Remains: Metrics, Statistics, and OsteoID. *Biology* **2022**, *11*, 25. [CrossRef] [PubMed]

biology

MDPI

Article

Adult Skeletal Age-at-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis

David Navega [1,2,*], Ernesto Costa [3] and Eugénia Cunha [1,2]

[1] Centre for Functional Ecology (CEF), Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; eugenia.m.cunha@inmlcf.mj.pt
[2] National Institute of Legal Medicine and Forensic Sciences, 3000-548 Coimbra, Portugal
[3] Centre for Informatics and Systems of the University of Coimbra (CISUC), Evolutionary and Complex Systems Group (ECOS), Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal; ernesto@dei.uc.pt
* Correspondence: dsnavega@gmail.com

Simple Summary: Age-at-death is of paramount importance in forensic analysis of skeletal remains. In addition to sex, stature, and population affinity, it constitutes baseline information in the identification process of deceased individuals. Despite its long tradition, in anthropological research age-at-death estimation poses many challenges and unanswered questions. It is undisputedly among the most difficult tasks of the forensic anthropologist and its results are often subject to a lackluster performance. In this study, we assessed computationally the efficiency of a holistic approach to skeletal age estimation based on a new proposal for macroscopic examination and the use of machine learning-based models for data analysis. Our results suggest that this approach is key for accurate and efficient age-at-death estimation based on skeletal remains analysis.

Abstract: Age-at-death assessment is a crucial step in the identification process of skeletal human remains. Nonetheless, in adult individuals this task is particularly difficult to achieve with reasonable accuracy due to high variability in the senescence processes. To improve the accuracy of age-at-estimation, in this work we propose a new method based on a multifactorial macroscopic analysis and deep random neural network models. A sample of 500 identified skeletons was used to establish a reference dataset (age-at-death: 19–101 years old, 250 males and 250 females). A total of 64 skeletal traits are covered in the proposed macroscopic technique. Age-at-death estimation is tackled from a function approximation perspective and a regression approach is used to infer both point and prediction interval estimates. Based on cross-validation and computational experiments, our results demonstrate that age estimation from skeletal remains can be accurately (~6 years mean absolute error) inferred across the entire adult age span and informative estimates and prediction intervals can be obtained for the elderly population. A novel software tool, DRNNAGE, was made available to the community.

Keywords: forensic anthropology; age-at-death estimation; machine learning; neural networks

Citation: Navega, D.; Costa, E.; Cunha, E. Adult Skeletal Age-at-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis. *Biology* **2022**, *11*, 532. https://doi.org/10.3390/biology11040532

Academic Editors: Stefan Bonn and Andrés Moya

Received: 11 February 2022
Accepted: 28 March 2022
Published: 30 March 2022

1. Introduction

Forensic anthropology (FA) has become a major component of forensic sciences. During recent decades, a profound change, a true paradigm change, has taken place and forensic anthropology has transformed itself into a discipline with its own theoretical and conceptual corpus and research agenda. It can be stated that the discipline and its attributes have evolved significantly. In fact, this evolution has been so marked and drastic that it can be argued that even some of the most experienced and long-term practicing anthropologists may have trouble conceptualizing and being fully proficient in the many areas now

covered by the discipline [1,2], or in even being able to foresee all possible interdisciplinary and technological developments. Nonetheless, biological profile estimation from human skeletal remains constitutes a pivotal task and inferring age-at-death, sex, stature, and population affinities is a fundamental step of the anthropological analysis in the context of the medico-legal identification process.

In the identification process of human remains, age-at-death is a major screening factor that helps reduce the universe of possible matches. Therefore, an estimate of this biological parameter is a normal request from police forces and judicial entities [3]. This process relies on a meticulous analysis of skeletal and dental structures with an association with chronological age-at-death. Although this is a topic in which significant research has been performed in recent decades, skeletal age estimation of adult remains continues to present many unanswered questions and challenges, especially for the elderly. Determining how to handle age estimation using multiple skeletal age-related traits remains among the problems most commonly identified for which a satisfactory solution has not yet been presented and research further is required [3–10]. Moreover, computational and statistical methods employed in the creation of age estimation techniques have been a topic of debate and contention [11–24].

The present work aims to lay a foundation to tackle some of the challenges of morphoscopic adult skeletal age estimation, especially in terms of its holistic or multifactorial aspect. Several authors argue in favor of multifactorial age estimation to obtain precise and accurate age estimates [9,16,25]. Nonetheless, multifactorial age estimation poses its own challenges and limitations, and is a topic with a clear lack of consensus [5,10]. Conceptually multifactorial age estimation can be argued as being the most effective approach for age estimation because morphological indicators display different age-related trajectories and have different underlying biological processes.

The symphyseal face of the pubic bone, for instance, has been systematically studied, ranging from the pioneering studies that established the morphological analysis of this skeletal marker as an age estimation technique, to modern fully computational frameworks for age estimation [26–34]. However, other skeletal markers and regions that can convey important age-related information, such as the degeneration of vertebral bodies, joint margins, or the roughening of muscle and tendon attachment sites, have received scarce attention as aging markers. The unimpressive accuracy and precision associated with the multiple iterations of pubic symphysis aging techniques, one of the most used and favored techniques for age estimation [5], underlines the idea that further developments and over-analysis of specific skeletal markers in isolation is not likely to result in substantial improvements over the state-of-art of adult age estimation, but rather a more comprehensive array of skeletal markers and features provide a more fertile ground for further developments [35,36].

A multifactorial morphoscopic approach to skeletal analysis does not solve, in itself, the many difficulties faced in the age-at-death assessment. In fact, if not correctly designed, this approach can become methodologically cumbersome from a data collection and analysis perspective. From an analytical and statistical perspective, collecting more data from the skeleton increases the chance of encountering issues of redundancy, multicollinearity, and a dimensionality that hinders the straightforward interpretability and pragmatic value of morphoscopic analysis. From a practical point of view, a more comprehensive analysis of the age-related skeletal features requires a higher level of expertise on how to collect the skeletal features. This issue is of great relevance for approaches that rely on morphoscopic analysis of the skeleton. Moreover, in forensic contexts it is common that the skeletal remains are somehow fragmentary or incomplete due to a multitude of taphonomic factors, which means that not all age-related traits will be available for every unidentified deceased. From a practitioner's perspective, this translates into the need for computational and software tools that can fit or train age-at-death estimation models on a case-by-case basis.

To cope with the difficulties and needs of multifactorial age estimation, novel methods and techniques can be developed by resorting to statistical and machine learning, data

science, and artificial intelligence tools and approaches. More than constantly evolving, machine learning, artificial intelligence and data science are ubiquitous, and have various successful applications within forensic anthropology in domains such as biological profiling or craniofacial identification [13,15,37–41].

This work aims to provide a new method, and its computational analysis, for multifactorial skeletal age-at-death estimation of adult humans supported by a machine learning approach based on a deep randomized neural network. This manuscript is in its essence methodological, presenting both a new macroscopic technique for skeletal analysis and a detailed explanation of a computational framework to obtain age-at-death estimates and model their uncertainty. New age-at-death estimation software, DRNNAGE, that translates the in silico key points of the work presented here into an actionable tool, was developed and is a major research product.

2. Materials and Methods

2.1. Dataset

2.1.1. Sampled Identified Skeletal Collections

To implement and pursue a computational analysis of the novel age-at-death estimation method proposed in this work, a reference dataset of 500 individuals was constructed. A total of 99 features were collected covering all key traditional age-related and other under-explored skeletal traits. Accounting for laterality, 64 unique traits can be analyzed from the axial and appendicular skeleton using the new macroscopic scoring method, whose rationale and details are described and explored in Section 2.2.

The 500 individuals were sampled from two identified skeletal collections hosted at the Department of Life Sciences at the University of Coimbra, Portugal—the Coimbra Identified Skeletal Collection (CISC) and the 21st Century Identified Skeletal Collection (XXI-ISC). The CISC consists of 505 individuals with age-at-death ranging from 7 to 96 years representing skeletons from the Cemitério da Conchada, that were born between 1817 and 1924 and died from 1904 to 1938 [42]. The XXI-ISC collection is currently composed of 302 skeletons of both sexes, mostly represented by elderly individuals. This collection represents Portuguese nationals who died between 1982 and 2012 and were exhumed between 1999 and 2016 from a main cemetery in Santarém. More details are found in [43,44]. Demographic parameters of the sampled individuals in our study are detailed in Table 1. All sampled individuals presented fully developed long bones. No individual was excluded due to pathology or taphonomy.

Table 1. Demographic characterization of reference data sampled from the CISC and XXI-ISC collections.

		CISC		XXI-ISC		Pooled Collections		Pooled Sex
		Female	**Male**	**Female**	**Male**	**Female**	**Male**	
	n	168	166	82	84	250	250	500
Age-at-Death	Mean	48.482	45.331	81.841	74.881	59.424	55.260	57.34
(AGE)	Std. Dev.	19.483	18.171	12.889	15.082	23.556	22.141	22.93
	Min.	19	19	38	25	19	19	19
	Max.	95	96	101	96	101	96	101
Year of Birth	Mean	1877.286	1879.994	1923.866	1930.560	1892.564	1896.984	1894.774
(YOB)	Std. Dev.	21.252	19.948	13.137	14.424	28.969	30.096	29.591
	Min.	1830	1836	1904	1908	1830	1836	1830
	Max.	1911	1917	1970	1982	1970	1982	1982
Year of Death	Mean	1925.768	1925.325	2005.707	2005.440	1951.988	1952.244	1952.116
(YOD)	Std. Dev.	6.597	7.343	3.707	3.919	38.051	38.452	38.214
	Min.	1910	1910	2000	1995	1910	1910	1910
	Max.	1936	1936	2012	2011	2012	2011	2012

The sampled reference dataset is composed of 250 male and 250 female individuals who died at the age of 19 to 101 years old (mean = 57.34, SD = 22.93). Age-at-death distribution is homogenous across the age span represented, with the exception of individuals over 95 years old (Figure 1). A homogenous and uniform age-at-death distribution is a simple

yet vital strategy to cope with the problem of age-mimicry [45] and to guarantee that the targeted age span is fully represented.

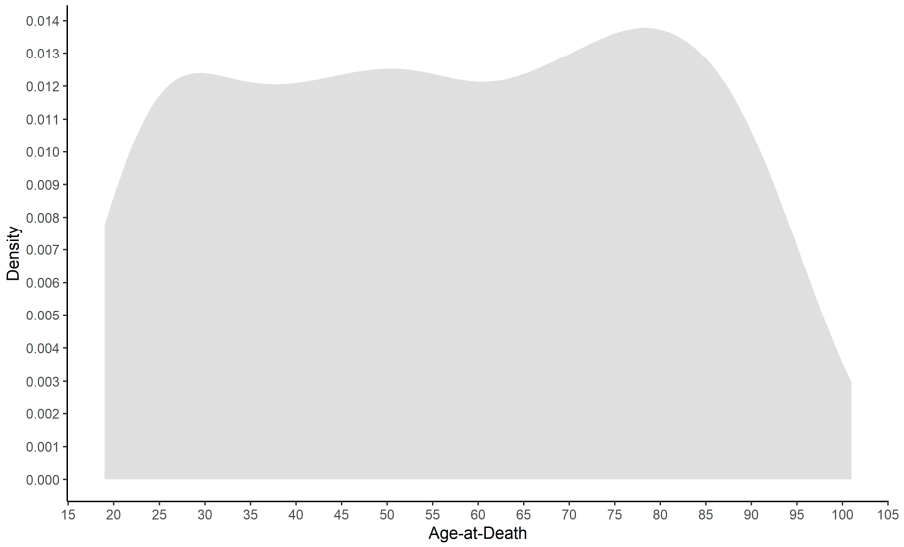


Figure 1. Pooled age-at-death distribution (KDE).

Sampled individuals were born between 1830 and 1982 and died between 1910 and 2012. Despite the large temporal frame represented, there is a continuum and a wide range over the age-at-death distribution that makes this sample particularly suited for age-related research.

2.1.2. Data Management and Processing

As previously mentioned, multifactorial age estimation poses many challenges that are mostly related to data management and processing. Two common problems that arise are redundancy and missing data. Redundancy is always involved when bilateral or paired data is collected. The human body is not fully symmetric; yet it is not expected that the left and right diverge drastically under normal conditions. Missing data in FA results mostly from taphonomic factors. To cope with redundancy and missing values, a strategy based on domain heuristics and imputation techniques was pursued. For bilateral traits, the left side was selected as the main source of data. If the left score for a given bilateral trait was missing, the right side was used as a surrogate value. Once this first heuristic was applied, the remaining missing values were imputed using a simple nearest neighbor (k = 1) procedure by substituting all missing value of given individual by the values of the nearest neighbor. Jaccard similarity on one-hot encoded data was used to compute the nearest matches. The followed procedure minimized redundancy and dimensionality by reducing the number of skeletal features from 99 to 64. A simple nearest neighbor with k = 1 according to Beretta and Santianello [46] is the preferred strategy to preserve the structure of a dataset. The authors demonstrated that more advanced algorithms reduced imputation error but introduced significant data distortion. To increase the volume and age-related variability of the data available, sexes were pooled. Although this choice seems arbitrary, it is important to note that, in FA, sex is usually estimated during casework. Pooled data models balance out the potential and pitfalls of sex-specific models and their mis-specifications.

Missing values represented 9.52% of the total entries of the data table when bilateral data were considered, and 6.89% when the domain heuristic described was first applied as a naïve imputation mechanism and strategy to handle bilateral data redundancy.

2.2. A Novel Technique for Macroscopic Age-At-Death Estimation

A key contribution of the present work to the topic of macroscopic skeletal age estimation in adults is the proposal of new scoring schemes for well-established and underexplored skeletal traits that can be used as biomarkers in age-at-death assessment. The development of a new scoring system emerged from the necessity for standardization of a data collection, and a generation mechanism that was more aligned with a multifactorial approach to age estimation and more suitable multivariate data analysis, while keeping in mind practical aspects such as observation error and ease of application.

The proposed morphoscopic method strives to be comprehensive and to incorporate features from as many skeletal elements as possible. Envisioning the whole skeleton as a biomarker for age estimation, it is more likely that the overall skeletal patterns exhibit a stronger and monotonic relationship with age-at-death, which is pivotal for accurate predictions. The rate and nature of overall skeletal changes also have a greater chance to be consistent across individuals since a holistic approach can encapsulate intra and interpersonal variation with greater finesse [35]. Analyzing multiple traits also offsets the intrinsic limitation to specific traits when analyzed on their own [47].

Following a component-based approach, up to 64 unique skeletal traits can be scored using the scheme outlined in the next subsections. The covered skeletal traits encode both developmental and degenerative aspects from different anatomical regions. Despite the large number of features analyzed in this proposal, all skeletal features are limited to morphological variables with no more than three classes or stages. Such specifications were established during the several iterations of the development and refinement of the system proposed, and by following guidelines from the literature. Shirley and Montes [48] empirically addressed the old methodological debate of phase versus component-based approach. Their study quantified the observation error of a phase and a component-based method, and the results suggests that a component-based approach offers a more objective scoring if the number of coding possibilities in each component does not exceed three levels of expression.

The following subsections provide a brief overview of the existing scoring methods for specific skeletal region or traits, the novel scoring schemes proposed in this work, and the rationale and difficulties faced during method development. Due to the constraints of space and manuscript presentation, full descriptions of the trait scoring systems developed in this study are provided in Tables S1–S15 of the Supplementary Material. The skeletal scoring systems are also embedded in the developed software (see Section 2.6.4).

2.2.1. Cranial and Palatine Suture Scoring

The scoring system used for the cranial and palatine sutures consists of a modification and binarization of the proposal by Boldsen et al. [19]. This system was selected because it incorporates much of the rationale of older methods for scoring ectocranial sutures (neurocranium) and the palatine sutures [49–56]. The simplification to a binary scoring system resulted from the difficulty during preliminary and training sessions to differentiate and consistently score the adjacent stage (i.e., open to juxtaposed or partially obliterated to punctuated). The scoring scheme described in Table S2 should be applied to nine sutural segments from the palatine, the sagittal, coronal, and lambdoid sutures (Table S1).

2.2.2. Vertebrae Development and Degeneration Scoring

The fusion of the bodies of the first and second sacral vertebrae is also part of the skeletal markers analyzed in the proposed protocol. This skeletal feature is one of the few developmental traits that persist through early adulthood. Its usefulness as an indicator to distinguish young adults was demonstrated by several researchers [57–59]. This trait was assessed with a binary scale described in Table S3. To incorporate both metamorphic and degenerative traits of the vertebral column, a three-stage scoring scheme was devised, building upon previous work from Snodgrass [60], Watanabe and Terazawa [61], and Albert et al. [62]. The first two methods focus on the degeneration and osteophyte formation on

the margins of the vertebral bodies, whereas the last work focuses on the development of the vertebral epiphyseal rings and body morphology. The proposed system, Table S4, applies to superior and inferior surfaces of the third to seven cervical vertebrae, the first to fifth lumbar vertebrae, and the superior surface of the first sacral vertebra. Table S5 lists all features analyzed in the axial skeleton (excluding sacral auricular surfaces).

2.2.3. Joint and Musculoskeletal Degeneration Scoring

Osteoarthrosis and entheseal changes have been traditionally analyzed in physical anthropology and bioarcheology as markers of health and biomechanical stress, and tentative indicators of physical activity patterns. According to Milner and Boldsen [35], who advocate a more detailed analysis of this type of skeletal marker, these features collectively contribute to an increase in accuracy and precision of age estimation. The authors base such an assertion on empirical evidence from an experience-based procedure where these types of skeletal traits were extensively used. Several reasons can be noted for why osteoarthrosis and entheseal changes have been overlooked or not systematically analyzed in the past as age markers. Broadly speaking, due to their degenerative nature and late onset, it is believed that they provide limited information, distinguishing only in a broad sense young from older individuals. More specifically, osteoarthrosis increases with age but has a complex and multifactorial etiology that hinders or masks its relationship with age-at-death. Entheseal changes have traditionally been assessed as musculoskeletal stress markers and as tentative clues to infer physical and occupational activity. This possible relation to activity can interfere in the expression and variation of entheseal morphology and affects its relationship with the aging process. However, recent and systematic studies conducted on identified skeletal collections show that age-at-death is one of the most relevant factors, or even the only one with statistical significance, in the expression of such skeletal traits [63–70].

Developing a scoring procedure for these features proved to be one of the most challenging aspects of method development. The difficulties faced were mostly related to the fact that analyzing joint and musculoskeletal degeneration involves many skeletal elements, which translate into high dimensionality of the collected data. This high dimensionality poses two major problems: increased chance of collinearity, which poses computational issues, and loss of pragmatic value. To tackle the high dimensionality and subsequent issues found when scoring joint and musculoskeletal degeneration, a new binary procedure was developed. The system retains the analysis of the type of traits evaluated in Buikstra and Ubelaker [71] and Henderson et al. [72] but simplifies the scoring to a simple absence or presence of degenerative traits as a whole for any particular anatomical structure. The generic binary scoring system both for joint and musculoskeletal degenerative changes are presented in Tables S7 and S8. The scoring system applies to five major anatomical complexes from the upper and lower limb: shoulder, elbow, hip, knee, and ankle (Table S6). To enhance the analysis of these traits we provide specific scoring descriptions for Stage 1 of some traits (Table S9).

2.2.4. Clavicle Sternal and Acromial Ends Scoring

The macroscopic analysis of the clavicle has a long standing in skeletal age estimation. Nonetheless, its focus has been mostly in the epiphyseal fusion of the sternal end [73–76]. Sternal epiphyseal fusion of the clavicle is a key trait to obtain precise age estimate in young adult individuals due to the late total development of this structure around the 30 s. Falys and Prangle [73] were the first to propose a method to score post-epiphyseal changes in the clavicle for age estimation purposes. The authors suggest a scoring system focused on surface topography, porosity, and marginal osteophyte formation, providing a regression model for age estimation. A new scoring scheme that integrates both developmental and degenerative changes in the sternal and acromial ends of the clavicle is proposed. A full description of the traits analyzed is available in Table S10.

2.2.5. First Rib Costal Face and Tubercle Scoring

The metamorphosis of the sternal end of the ribs emerged in the mid-1980s as a new age estimation technique. İşcan, Loth and colleagues described multiple morphologic features that characterize the metamorphosis of the sternal end of the ribs, with particular emphasis on the fourth rib costal face [77–80]. This approach proved to be an effective alternative to existing methods. Nonetheless, several disadvantages have been pointed out, such as the difficulty in identifying the fourth rib in disarticulated skeletal remains and the fact the morphology of the costal face is not the only component of the age-related changes in rib morphology. To address these problems, Kunos et al. [81] described a new age estimation method based on the metamorphosis of the costal face, head, and tubercle of the first rib. The first rib has the key advantage of having a morphology that is straightforward to individualize. DiGangi et al. [82] improved upon the work of Kunos et al. [81] and proposed a revised method for age estimation based on the costal face and tubercle morphology. A new scoring method is proposed in this study that build upon previous work by Kunos et al. and DiGangi et al. [81,82]. This new system simplifies the scoring of the costal face morphology to a three-stage coding and the morphology of the tubercle is evaluated in a binary fashion (Table S11).

2.2.6. Pubic Symphysis Scoring

The metamorphosis of pubic symphysis is the most popular osteological marker used in adult skeletal age estimation. The previous attention paid to this anatomical structure is not misplaced; however, the over-reliance on this indicator can be explained by the progressive metamorphic features that have enough expression variation to allow an exhaustive morphological description using different scoring schemes and different types of supporting materials such as casts. A simple component-based system was developed focused on the metamorphic and degenerative changes in three features of this structure: rim development, topography, and texture of the symphyseal face. These three components are assessed with a three-stage coding system emphasizing early metamorphic or development traits, such as the presence of billowing (a pattern of transverse ridges and furrows) and late degenerative traits, such as the flattening and erosion of the symphyseal face. A full description of the scoring system is given in Table S12. The proposed system is based on previous work by Todd [30,31] and Brooks and Suchey [26].

2.2.7. Sacral and Iliac Auricular Surfaces (Sacroiliac Joint) Scoring

The description of age-related changes in the sacro-iliac joint can be traced back to Sashin [83] and Schunke [84], but its usage as an age indicator its mostly due to the work of Lovejoy and colleagues [85] and Buckberry and Chamberlain [86] on the chronological metamorphosis of the iliac auricular surface, and the age estimation method by Passalacqua [59] based on metamorphic and degenerative changes in the sacrum.

To incorporate age-related features of the sacro-iliac joint, a two-component-based system was developed to assess textural and marginal changes in the sacral and iliac auricular surface. The iliac and sacral auricular surfaces undergo textural changes that are characterized by the transition from a smooth, finely grained surface to a granular, irregular and porotic surface. The margins that delimit the surface tend to manifest osteophytic activity as age progresses. Both the texture and margin features refer to the entire structure but very often the degenerative changes, in particular the margin, are more pronounced in specific areas such as the inferior and anterior apexes. Full features descriptions are given in Tables S13 and S14.

2.2.8. Acetabulum Scoring

Several age-related changes can be documented in the acetabulum and used for age estimation [87–94]. One key aspect of the acetabulum is the late onset of the age-related changes and its durability and resistance to taphonomic factors. To incorporate this skeletal element in our protocol, a three-stage scoring system for the changes occurring on the

rim, posterior horn, and acetabular fossa was developed. In the spirit of Calce [90], who simplified the method developed by Rissech et al. [91,92], the foundation of the scoring system presented in Table S15 is based on a simplification and adaptation of the method proposed by San-Millán et al. [87,95].

2.2.9. Scoring Reliability: Intra-Observer Error

To assess the reproducibility of this new proposed scoring system, 50 individuals were randomly selected and rescored on all possible traits (m = 99) by the first author. For bilateral traits, only the left side was used for further intra-observer reliability analysis (first author) to avoid issues that arise from non-independent ratings. Kendall's W [96] was computed as a concordance coefficient to assess consistency between scoring sessions. This metric ranges from 0 (no agreement) to 1 (perfect agreement).

2.3. Feature Analysis Via Sphering and Marginal Correlation Analysis

To assess the relationship of the analyzed traits with age-at-death, we inspected marginal correlation coefficients using Spearman's correlation coefficient (ρ) and Pearson's eta coefficient (η^2). In addition to these two coefficients, we also computed marginal correlations adjusted for inter-trait correlation following Zuber and Strimmer [97]. This technique aims to cope with the myopy of univariate feature selection methods by computing marginal correlations of decorrelated predictors with the target class. First, the data centered and scaled, and then transformed by applying a linear basis that enforces orthogonality among predictors while maintaining the maximum relationship with the original standardized predictors. After this transformation, also known as the Mahalanobis transform or sphering, the predictors covariance matrix is the identity matrix (no correlation). The authors called the adjusted marginal correlations CAR scores and proved that ranking based on these quantities provides a fast and optimal procedure for feature ranking and selection. We suggest [97,98] as primers on feature selection and data sphering based on this approach.

2.4. Randomized Neural Networks: Theory and Implementation

From a computational perspective, age-at-death estimation can be viewed as a function approximation problem, $y = f^*(x)$, and constitutes one of the core reasons why artificial neural networks were chosen as the predictive technique in this work. In age-at-death estimation, $y = f^*(x)$ maps the input skeletal traits (x) to an age-at-death (y). ANNs are function approximation machines that define the mapping $y = f(x;\theta)$, where θ are the parameters or network weights that result in the best approximation [99].

Artificial neural networks are a class of connectionist, biologically inspired computational models that enable learning from data for a multitude of tasks, such as classification, regression, representation learning, and data compression and generation. ANNs are, in a broad sense the result of two components: architectural design—that is how many layers and neurons comprise the network; and an optimization algorithm—how the parameters of the network are learnt.

In its basic implementation, an ANN is composed of three layers: an input layer, a hidden layer, and an output layer. Two sets of weights are embedded in the network structure: one connecting the inputs to the hidden layer and the other connecting the hidden layer to the output layer. In a neural network, the input is transferred to the hidden layer by means of a non-linear activation function. An activation function and the set of weights define a node of the hidden layer. Such nodes are also known as artificial neurons. An artificial neuron, the key component of an ANN, is a mathematical operator in the form of:

$$h(x) = g(\sum_{i=1}^{p} x_i \omega_i + b) \quad (1)$$

where $g()$ is an activation or transfer function, x_i and ω_i are the i-th components of the input, and the weight vector b is the neuron bias. Artificial neurons are, in essence, non-

linear functions with learnable parameters, which ultimately expand the ANN model representational capacity to be able to approximate any output function.

A key aspect of ANN is their flexibility and modularity, which due to their capability can be applied to a vast array of heterogeneous data types and domains. The explosion in the availability and capacity to store and analyze data in the form of images, video, audio, and unstructured text has led to the development of novel ANN training algorithms and architectures, and a transition from shallow (single hidden layer) to deep (multi-layer) networks. It is important to note that not all ANNs are formulated and trained in the same manner. There are specialized architectures to tackle; for instance, data in the form of images that make use of computational operations, such as convolutions and pooling. However, a transversal aspect of modern ANNs is their use of gradient-based learning algorithms, where the parameters of a network are iteratively fine-tuned. Gradient-based learning enables end-to-end training and state-of-the-art performance in many complex tasks, but it is costly and requires considerable amounts of technical knowledge to leverage an ANN to its full potential.

A counterintuitive, yet highly efficient, approach to the training of ANN models is to randomly assign and fix a subset of parameters (i.e., hidden weights) of the network and recast the optimization component to a simpler least squares estimation problem [100,101]. In the context of ANNs, randomization as an intrinsic mechanism of model learning can be traced back to late 1980s and early 1990s, with the proposal of randomized radial basis functions network (RBF) and the random vector functional link network (RVFL) models [102–106]. However, the recent interest in randomized algorithms for training feed-forward neural networks can be attributed to the re-emergence of this approach in the guise of the controversial extreme learning machine (ELM) algorithm [107–110]. According to [111], there is no need to rename this strategy for training neural networks, since all key elements have been previously proposed [102–106], and some of the minor changes introduced by the ELM algorithm, such as the omission of direct links between the input and output layer—present in the RVFL network—can have a deleterious effect in performance. Nonetheless, the ELM algorithm acted as a foundation for many innovations in the field of randomized artificial neural networks (RANNs), such as the development of highly efficient algorithms to compute and cross-validate the output layer analytically [112,113], and its evolution from a framework restricted to shallow networks to a set of techniques and algorithms capable of deep, multi-layered network architectures [114–118].

2.4.1. Efficient Training and Regularization in Randomized Neural Networks

In randomized neural networks, the elements of ω_i, the hidden layer weights, are randomly generated from a suitable probability distribution and are not optimized. Only the output weights are learned from data by solving a least squares estimation (LSE) problem expressed as:

$$\beta = H^{\dagger} Y \tag{2}$$

where β are the output layer weights, $H^{\dagger}$ is the Moore–Penrose pseudo-inverse of the matrix H, which defines the hidden layer, and Y is a column vector storing the network target output, in our case, age-at-death. $H^{\dagger}$ can be computed using several methods; a common approach is through orthogonal projection using Equation (3):

$$H^{\dagger} = \left(H^T H\right)^{-1} H^T \tag{3}$$

From Equations (2) and (3), it is trivial to show that the use of this algorithm yields an age estimate as $\hat{Y} = H\beta$, and that the output layer is in fact an ordinary least squares linear regression built on the non-linear feature mapping induced by the hidden layer of the neural network.

It has been noted [119] that one can keep the algorithmic simplicity of the least squares solution, while improving its performance and generalization capability by adding a penalty to the output weights. Such a penalty, C, stabilizes the inversion of matrix H and

shrinks the coefficients of the output layer towards zero; smaller coefficients lead to smaller error rates on unseen data. Imposing such a constraint on the output weights is a process known as shrinkage or regularization, which in the neural network literature is also named weight decay. This type of regularization is also referred as L2-norm regularization or Tikhonov regularization.

The solution of a regularized RANN is obtained by fitting a ridge regression model [120] as the output layer. The ridge solution, β_{ridge}, is obtained by substituting Equation (3) as follows:

$$H^{\dagger} = (H^T H + \frac{I}{C})^{-1} H^T \tag{4}$$

I refers to the identity matrix with dimensions matching $H^T H$. Regularization is of paramount importance when training a randomized neural network for age estimation. The solution of the network is obtained by minimizing the squared error as the objective function. LSE-based neural networks lead to unbiased solutions but with high variance if not properly regularized due to the randomness of the initialization [112]. Regularization shrinks the size of the output coefficients towards zero, which is consistent with the theory that smaller weights result in better generalization of neural networks [121,122].

Since the output layer in a RANN is solved as a least squares estimation problem, fortunately, there exist highly efficient, analytical, and closed formulations to assess the leave-one-out (LOO) error, as shown by Shao and Er [112] using Allen's [123] Prediction Sum of Squares (PRESS) statistic:

$$E_{LOO} = \frac{1}{n} \sum_{i=1}^{n} (\frac{y_i - \hat{y}_i}{1 - h_{ii}})^2 \tag{5}$$

where h_{ii} is the i-th diagonal element of the hat or projection matrix, which is the matrix that maps the hidden layer parameters to the predicted values of the network, in our case age-at-death. Shao and Er [112] have demonstrated that computing the projection matrix of the network and finding the optimal regularization parameter, C, under leave-one-out cross-validation (LOO-CV), can be achieved with computational efficiency by performing a singular value decomposition (SVD) of the hidden layer, which, given such an operation, is written as $H = U\Sigma V^T$. Using SVD, the network estimate can be written as:

$$\begin{aligned} \hat{Y} &= H\beta \\ \hat{Y} &= H(H^T H + \tfrac{I}{C})^{-1} H^T Y \\ \hat{Y} &= U(\Sigma^T \Sigma + \tfrac{I}{C})^{-1} \Sigma^T U^T Y \end{aligned} \tag{6}$$

where $U\left(\Sigma^T \Sigma + \frac{I}{C}\right)^{-1} \Sigma^T U^T$ is the projection matrix and it can be noted that only $\left(\Sigma^T \Sigma + \frac{I}{C}\right)^{-1} \Sigma^T$ affects the projection matrix for different values of C. Σ is a diagonal matrix whose element are expressed as $\phi_i = \frac{\sigma_{ii}^2}{\sigma_{ii}^2 + \frac{1}{C}}$, where σ_{ii} is the i-th singular value from the decomposition of H. SVD makes the regularization of the neural network highly efficient because the diagonal of the projection matrix, which is needed to calculate the LOO error using Equation (6), can be obtained from the following Hadamard products (matrix element-wise multiplication):

$$\gamma = U \circ \Gamma^T = U \circ (\Theta \circ U^T) \tag{7}$$

where $\Theta = \left(\Sigma^T\Sigma + \frac{I}{C}\right)^{-1}\Sigma^T$. The diagonal elements of the projection matrix, h_{ii}, can be obtained by performing a column-wise sum of the elements of γ. The LOO predictions of the network can be obtained analytically as follows:

$$\hat{y}_i = \frac{y_i - f(x_i)}{1 - hat_{ii}} \tag{8}$$

In addition to this highly efficient computational strategy to train a randomized neural network, data standardization and the addition of Gaussian noise to several of the components of the network can also improve performance and accuracy.

2.4.2. From Shallow to Deep Randomized Neural Networks

The mathematical and network formulation presented above pertain to a randomized weights single layer network architecture. Navega and Cunha [124] introduced this model in skeletal age estimation in the formulation of the ELM network (no direct links in the network) and applied it to several traits of the sacroiliac joint. However, several authors proposed different techniques to extend the RANN to deeper architectures [114–118]. To increase the deepness of the network, one can resort to fully randomized approaches or use autoencoding strategies and stack multiple autoencoding RANNs to build a multi-layer network. In this work, due to its simplicity, we follow the proposal of Shi et al. [118] to train deep randomized network models (DRNNs). Following the authors, the first layer of the network is defined as:

$$H^{(1)} = g(XW^{(1)}) \tag{9}$$

where X is the input matrix, in our case skeletal traits. Every subsequent layer ($j > 1$) is defined as:

$$H^{(j)} = g(H^{(j-1)}W^{(j)}) \tag{10}$$

where $H^{(j-1)}$ is the previous layer. One can also allow connections from the input to all hidden layers and define the hidden layer as:

$$H^{(j)} = g([H^{(j-1)}\ X]W^{(j)}) \tag{11}$$

where W^1 and W^j are the weight matrices between the input-first hidden layer and the inter-hidden layers, respectively. These matrices are randomly assigned and held fixed during the training. The input to output layer is then defined as:

$$D = \left[H^{(1)}\ H^{(2)}\ \ldots\ H^{(j-1)}\ H^{(j)}\ X\right] \tag{12}$$

The design of the deep network is very similar to that of a shallow RANN, and it can be easily seen that the input to output layer consists of non-linear features induced by the hidden layers concatenated to the original input of the network. When the input is reused directly in the output layer, the network is classified as a network with direct link or skip layers. As mentioned above, this is the key difference between ELM and RFVL networks.

2.4.3. Deep Random Neural Networks as Implicit Ensemble Models

One key advantage of the randomized approach used in this study is that it can enable implicit neural ensemble models [118]. Rather than applying Equation (2) once to solve the output layer weights (solution), Equation (2) can be re-used along the depth of the network for each $H^{(j)}$ computed from Equations (9) or (10), and obtain an intermediate age-at-death estimate. The final age-at-death estimate can be then obtained by averaging all estimates along the network depth. This feature stabilizes the predictions and offers a different mechanism to train an ensemble model other than training each model independently.

2.5. Regression Uncertainty Modeling and Prediction Intervals

The approach followed in this work relies heavily on regression. In Sections 2.4.1 and 2.4.2, we presented the foundation for mathematical age-at-death prediction using RANN models as a regression task. However, we focused only on how point estimates can be obtained, that is, the conditional expectation of age-at-death given a specific skeletal pattern of an individual. Mapping the uncertainty of the point estimate is essential in forensic anthropology, which means that a predictive interval for a preset confidence level should also be part of the analysis and the subsequent report.

In the current work, we follow a simple and generic approach based on modeling the conditional variance associated with each point estimate (network prediction). We recast the prediction interval construction as a regression problem and, using LOO network predictions, we build a regression uncertainty model (RUM) by regressing absolute residuals on predicted age-at-death. We then scale the predicted residual by 1.2533 to obtain a standard deviation associated with each age estimate. The scaling factor is the ratio of the standard deviation to the absolute deviation [125,126]. Assuming normality of the variance around each point estimate, the prediction interval associated with an ANN model is given by the quantiles of a Gaussian or truncated Gaussian parameterized with the conditional mean and standard deviation inferred from the ANN and its associated RUM. The key advantage of this approach is its simplicity compared to likelihood methods [15–17,20,23,127–129] or conformal prediction theory, as in [113,124,130]. In addition to the numerical interval, this approach also allows visualization, as illustrated by Figure 2.

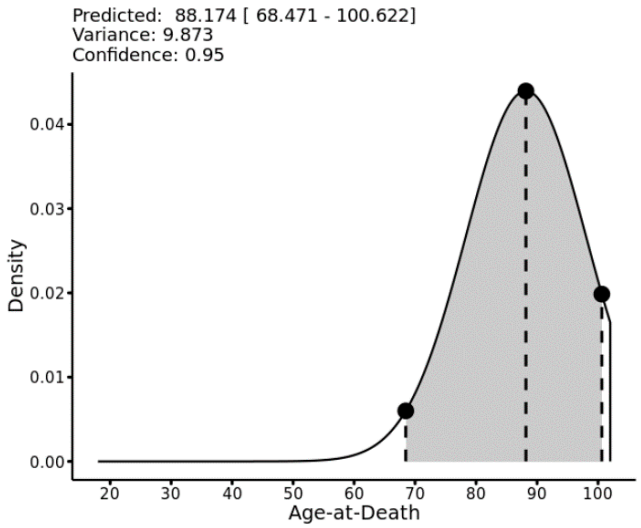


Figure 2. Prediction interval visualization using a (truncated) Gaussian uncertainty model.

2.6. Computational Analysis: Design, Parameterization, Metrics, and Software

2.6.1. Experimental Design

To assess the performance of DRNN and Gaussian RUM models in multifactorial age estimation from macroscopic skeletal traits we followed a simple template for robust metric assessment based on a resampling Monte Carlo cross-validation (MCCV) scheme. This works as follows: for a given iteration of the scheme, split the dataset into disjoint train and test partitions. Using the training partition, fit a DRNN and RUM models by making use of Equations (5)–(7) to optimize the regularization parameter C and obtain leave-one-out predictions. C is optimized as 2^x with $x \in \{-6, -4, \ldots, 12\}$. With the trained DRNN and RUM models, we predict the age-at-death of the testing sample/partition and compute the MCCV performance metrics. For a given set of skeletal traits, this procedure is repeated 1000 times (B = 1000). The training partition is set as 80% of the total data (400 of 500) and the test partition as the remaining (100 of 500). This sampling procedure

was performed without replacement. The core of our computational analysis is organized in two experiments, from now on referred to as experiments A and B:

(A) The first experiment we conducted was designed to provide a baseline of the accuracy obtained by fitting DRNN models to blocks of traits that have standard or traditional analytical framing. For instance, we fitted models to different anatomical complexes or sets of traits that mimic existing aging standards, i.e., a model for the sutures or the pubis symphysis.

(B) Our second computational experiment consisted of simulated different proportions of available traits from 90% to 10%. The objective of this experiment was to assess model performance in a more realistic scenario where the forensic anthropologist has skeletal traits available on a case-by-case basis.

In both experiments we computed 95% predictive intervals (95% PI) by setting the uncertainty of parameter $\sigma = 0.05$.

2.6.2. Network Parameterization

A key aspect of any ANN model is its architecture, that is, how many neurons (or nodes) and layers comprise the network. To leverage the full potential of the DRNN, and to maximize its training speed and efficiency, rather than search for the optimal architecture, we developed a simple heuristic based on the work of Lappas [131]. The author demonstrated that the size of a single layer perceptron can be estimated from the number of samples available. Using his work as a foundation, we propose the following heuristics for setting the architecture of a DRNN. The width, size, or number of neurons of each layer was set as:

$$S = 2^{\lfloor \log_2 (8(\sqrt{2^k}/k)) \rfloor}, \; k = \log_2(n) \tag{13}$$

where n is the number of samples. The depth or number of layers was set as:

$$L = 2^{\lfloor \log_2(k) \rfloor}, \; k = \log_2(n) \tag{14}$$

Following Equations (13) and (14) as a simple heuristic allows us to have predictable, parsimonious network architectures. In this way, the network allows many computing units for randomized feature extraction distributed over several layers without incurring overparameterization. This heuristic also leverages the simplicity of training a deep neural network using the same mechanisms of a shallow one, while exploiting an implicit ensemble framework (Section 2.4.3). For our experiments, applying the described heuristic defines the network architecture with a rectangular topology comprising eight layers of 32 neurons each, for a total of 256 randomized units.

DRNNs are computationally cheap nonlinear models built by combining regularized linear regression with nonlinear features obtained by using an activation function, $g(.)$, with random weights. In this work, we used the rectified linear unit (ReLU) as the nonlinearity of the networks. The ReLU is defined as $g(z, w) = \max(0, zw)$, where z and w are the layer input and random weight matrices. Since the regularization process involved in the training process described in this work is not scale invariant, during network training normalization by mean centering and variance scaling, Equation (6) was performed on the matrices X, XW, H, and Y. The output of the network was later rescaled before computation of the performance metrics.

ANN architecture selection and design is a non-trivial task often performed through very expensive and complex computational strategies and procedures. The heuristic used and architecture selected in this work emerged from trial-and-error experimentation during the development of the *rwnnet* software package (see Section 2.6.4). This parameterization leverages the benefits and key features of randomized neural networks—fast training and prediction with minimum technical knowledge, given that the model is fully described through linear algebra and matrix operations.

2.6.3. Performance Metrics

In our analysis, we evaluate four parameters that any model used in regression task should have, especially one used for age estimation. An age-at-death prediction model—regardless of its underlying mathematical algorithm—should be accurate, unbiased, valid, and efficient. Accuracy refers to the ability of the model of the model to predict age with minimal error. The most straightforward metric to assess this parameter is the mean absolute error (*MAE*) computed as:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{15}$$

where y_i and $\hat{y}_i$ are the known and predicted values, respectively, and n is the number of evaluated samples.

A model should be unbiased, that is, free of systematic error. A typical pattern of bias or systematic error in age estimation models is the over-estimation of young individuals and under-estimation of the elderly. A robust and comprehensive way to assess bias ($\hat{\beta}_e$) is by computing the slope of the regression line of the residuals, $e_i = y_i - \hat{y}_y$, on known values. When minimal to no bias is presented, this value should be close to zero. A positive slope suggests a systematic bias, such as the one describe previously. Bias is computed as:

$$\hat{\beta}_e = \frac{\sum (y_i - \overline{y})(e_i - \overline{e})}{\sum (y_i - \overline{y})^2} \tag{16}$$

where $\overline{y}$ and $\overline{e}$ are the means of the known and residual values.

The validity of model, in the context of our study, refers to the ability of a model to contain the known age within the predictive interval and within a reasonable margin close to the nominal uncertainty level allowed. For instance, for an uncertainty level (alpha) of 0.05 (or 5%) we expect that the coverage of the correct proportion of individuals within the predictive interval is close to 0.95 (or 95%). As a validity measure, we compute:

$$P(\alpha) = \frac{\sum_{i=1}^{n} \delta(y_i, l_i, u_i)}{n} \tag{17}$$

where $\delta(y_i, l_i, u_i)$ is an indicator function with $\delta(y_i, l_i, u_i) = 1$, if $y_i \geq l_i \wedge y_i \leq u_i$ and $\delta(y_i, l_i, u_i) = 0$, and l_i and u_i are the values of the lower and upper ends of the predictive interval, respectively.

Finally, a model should thrive to be efficient. Efficiency in this context refers to the width or range of the prediction intervals associated with the regression uncertainty model. A method or model is efficient when it outputs the narrowest predictive interval possible while also maintaining its validity. We compute our measure of efficiency as follows:

$$\text{PIW} = Q(u - l,\ \tau), with\, \tau\ \in \{0.5;\ 0.025;\ 0.975\} \tag{18}$$

where $Q(.)$ is a quantile function and τ a given quantile. One can see that we compute the median of the predictive interval width and its associated 95% confidence interval (quantile-base).

2.6.4. Software

All computational work was performed using the R and C++ programming languages with all key software components written by the first author. To perform this work, the rwnnet, rumr, rmar, and lsmr packages were used. These packages are available from the respective repositories of the GitHub profile of the first author, https://github.com/dsnavega (accessed on 18 March 2022).

Novel software, DRNNAGE, that operationalizes age-at-death estimation following the macroscopic and computational techniques described in this work, was also devel-

oped and is live as a web application at https://osteomics.com/DRNNAGE (accessed on 18 March 2022).); its source is available at https://github.com/dsnavega/DRNNAGE (accessed on 18 March 2022).). In its current state, we strongly recommend that end users approach their analysis using only default parameters. All problems detected and suggestions should be directed to the corresponding author.

3. Results

3.1. Intra-Observer Scoring Error

Overall, the new proposed macroscopic scoring technique presented high intra-observer consistency based on the results on Kendall's W concordance coefficient [96]. With the exception of RD01 and FM01, 0.751 and 0.716, respectively, all skeletal traits presented a concordance coefficient higher than 0.800. The global average of this coefficient was 0.907. All traits presented a statistically significant concordance between scoring obtained by the first author in two different sessions. The high concordance observed can be explained by the simplicity of the scoring systems used with the large number of traits that were binary coded. Further inter- and intra-observer error analysis is required by an independent third party, due to the nature of the methods employed.

3.2. Marginal Correlation Analysis

Marginal correlation analysis showed that all traits have a statistically significant relationship with age-at-death. The cranial sutures showed the lowest marginal correlation (ρ: 0.297–0.519, η^2: 0.088–0.249), with palatine sutures explaining less than 10% of the variation in observed age-at-death. The axial traits—cervical and lumbar vertebrae—exhibited a moderate to strong monotonic relationship and explained variation with age-at-death (ρ: 0.794–0.845, η^2: 0.639–0.725). A similar correlation and explained variation pattern were observed for the clavicle traits (ρ: 0.710–0.851, η^2: 0.507–0.729), first rib traits (ρ: 0.763–0.776, η^2: 0.590–0.607), iliac auricular surface traits (ρ: 0.731–0.789, η^2: 0.539–0.631), and the acetabular traits (ρ: 0.782–0.818, η^2: 0.625–0.674). A slightly lower marginal correlation was observed for the pubic symphysis traits (ρ: 0.711–0.731, η^2: 0.523–0.549) and sacral auricular surface traits (ρ: 0.632–0.704, η^2: 0.398–0.499). Traits from the upper and lower limbs presented a wider range of correlation (ρ: 0.380–0.789, η^2: 0.145–0.628). When analyzed in the context of feature ranking based on marginal correlations adjusted for inter-trait correlation (CAR scores), the suture traits score was among the worst predictors and its decorrelated components showed no statistically significant relationship with age-at-death. The several appendicular degenerative traits—HM04, UL01, RD01, FM01, FM02, and TB01—also showed no statistically significant correlation when assessed on a Mahalanobis transformed space. Ranking based on CAR scores showed that the top-ranking traits came from all anatomical regions rather than a specific indicator.

3.3. Computational Model Assessment

Results from the two in silico experiments performed to assess DRNN models in age-at-death estimation are reported in Tables 2–5. Models based solely on the cranial sutures exhibited the worst performance among all models produced, having a median MAE of 15.300 (Table 2) and a median predictive interval width (PIW) of 68.144 years, which renders the cranial sutures an inaccurate and inefficient set of traits.

Table 2. Monte Carlo cross-validation metrics for DRNN models built on pre-specified skeletal traits sets.

		Accuracy	Bias	Validity	Efficiency		
Traits		**MAE**	$\hat{\beta}_e$	$P(\alpha)$	**PIW**	**PIW 95% CI**	
Sutures (m = 9)	Median	15.300	0.656	0.950	68.144	51.699	69.759
	95% CI	13.586	0.590	0.900	66.054	46.361	68.312
		17.206	0.732	0.990	69.741	55.776	70.963
Axial (m = 16)	Median	8.185	0.198	0.960	38.754	33.732	40.842
	95% CI	7.365	0.137	0.920	37.102	32.272	39.215
		9.139	0.260	0.990	40.091	35.029	42.191
Appendicular (m = 23)	Median	7.583	0.167	0.960	37.378	29.109	39.541
	95% CI	6.678	0.103	0.910	35.412	27.613	38.014
		8.523	0.231	0.990	39.079	30.399	41.061
Clavicle (m = 2)	Median	8.949	0.244	0.960	49.234	17.354	51.610
	95% CI	7.798	0.169	0.920	39.064	15.981	49.962
		10.192	0.307	0.990	52.688	18.617	53.098
First Rib (m = 2)	Median	9.500	0.277	0.950	48.936	24.334	49.637
	95% CI	8.138	0.204	0.900	46.879	22.499	47.687
		10.831	0.351	0.990	50.903	26.078	51.533
Pubic symphysis (m = 3)	Median	10.897	0.370	0.940	51.210	26.905	56.954
	95% CI	9.371	0.280	0.870	48.688	24.520	54.799
		12.542	0.459	0.980	55.558	29.058	58.802
Sacroiliac complex (m = 6)	Median	8.523	0.223	0.950	44.668	20.378	47.969
	95% CI	7.380	0.145	0.890	39.350	18.596	46.017
		9.742	0.288	0.990	47.547	21.915	49.720
Acetabulum (m = 3)	Median	8.886	0.229	0.970	42.978	31.727	45.742
	95% CI	7.758	0.162	0.920	41.201	29.897	43.891
		10.006	0.287	1.000	44.509	33.240	47.304
Degenerative traits (m = 39)	Median	6.962	0.147	0.970	33.732	28.882	35.122
	95% CI	6.084	0.085	0.920	32.460	27.570	33.488
		7.814	0.200	1.000	34.935	30.019	36.656
Standard traits (m = 16)	Median	6.609	0.147	0.950	34.245	12.927	41.087
	95% CI	5.561	0.087	0.890	29.701	11.833	39.097
		7.598	0.202	0.990	37.857	14.169	42.833
All (m = 64)	Median	5.925	0.117	0.950	30.010	15.631	36.081
	95% CI	5.101	0.060	0.900	26.817	14.464	34.612
		6.728	0.170	0.990	33.191	16.811	37.515

Table 3. Leave-one-out cross-validation metrics for DRNN models built on pre-specified skeletal traits sets.

		Accuracy	Bias	Validity	Efficiency		
Traits		**MAE**	$\hat{\beta}_e$	$P(\alpha)$	**PIW**	**PIW 95% CI**	
Sutures (m = 9)	Median	15.245	0.655	0.953	68.120	51.782	69.796
	95% CI	14.683	0.616	0.940	66.377	46.429	68.371
		15.751	0.692	0.963	69.708	55.878	70.996
Axial (m = 16)	Median	8.156	0.200	0.960	38.825	33.594	40.881
	95% CI	7.896	0.184	0.953	37.468	32.131	39.279
		8.394	0.213	0.968	39.872	34.902	42.234
Appendicular (m = 23)	Median	7.557	0.169	0.960	37.534	29.035	39.599
	95% CI	7.278	0.155	0.948	35.996	27.542	38.082
		7.823	0.184	0.970	38.920	30.319	41.109
Clavicle (m = 2)	Median	8.943	0.245	0.963	49.216	17.336	51.768
	95% CI	8.606	0.228	0.953	47.184	15.969	50.112
		9.248	0.263	0.970	51.238	18.597	53.252

Table 3. *Cont.*

		Accuracy	Bias	Validity	Efficiency		
Traits		**MAE**	$\hat{\beta}_e$	$P(\alpha)$	**PIW**	**PIW 95% CI**	
First Rib (m = 2)	Median	9.409	0.275	0.950	48.897	24.356	49.811
	95% CI	9.067	0.255	0.938	47.036	22.502	47.862
		9.751	0.296	0.960	50.829	26.102	51.724
Pubic symphysis (m = 3)	Median	10.898	0.370	0.932	51.113	27.029	57.040
	95% CI	10.436	0.343	0.922	48.668	24.616	54.949
		11.315	0.398	0.945	53.003	29.217	58.909
Sacroiliac complex (m = 6)	Median	8.438	0.220	0.950	44.765	20.350	48.037
	95% CI	8.075	0.200	0.940	42.461	18.607	46.091
		8.741	0.239	0.960	46.755	21.893	49.800
Acetabulum (m = 3)	Median	8.833	0.229	0.965	43.051	31.541	45.832
	95% CI	8.490	0.210	0.955	41.302	29.726	43.995
		9.116	0.247	0.975	44.535	33.054	47.395
Degenerative traits (m = 39)	Median	6.929	0.147	0.963	33.744	28.816	35.194
	95% CI	6.694	0.133	0.953	32.530	27.499	33.566
		7.154	0.157	0.973	34.829	29.946	36.715
Standard traits (m = 16)	Median	6.561	0.145	0.948	34.283	12.952	41.170
	95% CI	6.277	0.132	0.935	32.464	11.853	39.222
		6.855	0.157	0.960	36.027	14.122	42.921
All (m = 64)	Median	5.899	0.118	0.950	30.057	15.558	36.141
	95% CI	5.677	0.110	0.940	28.758	14.403	34.644
		6.121	0.127	0.963	31.485	16.668	37.620

Table 4. Monte Carlo cross-validation metrics for DRNN models built on different fractions of available skeletal traits.

		Accuracy	Bias	Validity	Efficiency		
Available Traits %		**MAE**	$\hat{\beta}_e$	$P(\alpha)$	**PIW**	**PIW 95% CI**	
90% (m ≈ 57)	Median	5.964	0.120	0.950	30.354	15.851	36.215
	95% CI	5.136	0.062	0.900	27.067	14.466	34.554
		6.773	0.169	0.990	33.422	18.081	37.705
80% (m ≈ 51)	Median	6.026	0.121	0.950	30.498	16.004	36.261
	95% CI	5.211	0.061	0.900	27.183	14.213	34.498
		6.851	0.172	0.990	33.584	18.492	37.902
70% (m ≈ 44)	Median	6.072	0.125	0.950	30.805	16.206	36.454
	95% CI	5.152	0.062	0.900	27.528	14.001	34.600
		6.924	0.180	0.990	34.004	19.666	38.405
60% (m ≈ 38)	Median	6.131	0.125	0.950	30.964	16.352	36.649
	95% CI	5.316	0.065	0.900	27.513	13.893	34.672
		7.049	0.179	0.990	34.320	20.532	38.692
50% (m ≈ 32)	Median	6.237	0.129	0.950	31.479	16.717	36.969
	95% CI	5.293	0.064	0.900	27.820	13.757	34.930
		7.180	0.179	0.990	34.854	22.119	39.250
40% (m ≈ 25)	Median	6.360	0.134	0.950	32.125	17.165	37.429
	95% CI	5.441	0.074	0.900	28.500	13.910	35.075
		7.380	0.193	0.990	35.636	23.292	40.166
30% (m ≈ 19)	Median	6.570	0.140	0.950	33.163	17.933	38.137
	95% CI	5.565	0.075	0.900	29.036	13.905	35.393
		7.651	0.201	0.990	36.916	25.407	40.861
20% (m ≈ 12)	Median	6.951	0.153	0.950	35.263	19.946	39.694
	95% CI	5.857	0.086	0.900	31.082	14.074	36.427
		8.139	0.218	0.990	39.625	28.892	43.619
10% (m ≈ 6)	Median	8.026	0.196	0.950	39.618	26.914	43.025
	95% CI	6.592	0.119	0.900	34.681	15.495	38.368
		9.683	0.276	0.990	46.043	34.276	49.479

Table 5. Leave-one-out cross-validation metrics for DRNN models built on different fractions of available skeletal traits.

		Accuracy	Bias	Validity	Efficiency		
Available Traits %		**MAE**	$\hat{\beta}_e$	$P(\alpha)$	**PIW**	**PIW 95% CI**	
90% (m ≈ 57)	Median	5.942	0.121	0.953	30.276	15.745	36.278
	95% CI	5.699	0.110	0.940	28.748	14.339	34.599
		6.198	0.131	0.965	31.797	18.048	37.772
80% (m ≈ 51)	Median	5.970	0.122	0.953	30.476	15.941	36.332
	95% CI	5.702	0.108	0.940	28.860	14.162	34.574
		6.235	0.132	0.965	31.963	18.470	37.938
70% (m ≈ 44)	Median	6.028	0.124	0.953	30.711	16.182	36.518
	95% CI	5.737	0.108	0.938	28.960	14.013	34.697
		6.376	0.137	0.965	32.583	19.643	38.435
60% (m ≈ 38)	Median	6.078	0.125	0.953	30.975	16.342	36.716
	95% CI	5.768	0.108	0.938	29.070	13.872	34.756
		6.441	0.140	0.965	33.017	20.569	38.732
50% (m ≈ 32)	Median	6.173	0.128	0.953	31.502	16.684	37.040
	95% CI	5.819	0.111	0.938	29.410	13.724	34.989
		6.648	0.146	0.968	33.900	22.110	39.305
40% (m ≈ 25)	Median	6.305	0.132	0.953	32.146	17.153	37.511
	95% CI	5.903	0.114	0.935	29.839	13.905	35.130
		6.797	0.153	0.968	34.565	23.287	40.214
30% (m ≈ 19)	Median	6.501	0.138	0.953	33.097	17.923	38.203
	95% CI	6.046	0.118	0.935	30.583	13.899	35.468
		7.096	0.163	0.965	35.986	25.377	40.943
20% (m ≈ 12)	Median	6.957	0.154	0.953	35.321	19.986	39.742
	95% CI	6.316	0.127	0.935	32.096	14.117	36.479
		7.674	0.184	0.968	38.931	28.768	43.707
10% (m ≈ 6)	Median	7.952	0.192	0.955	39.733	26.846	43.076
	95% CI	6.968	0.154	0.940	35.229	15.515	38.419
		9.214	0.256	0.973	46.437	34.087	49.551

Modeling based on specific anatomical regions resulted in a DRNN with a median MAE ranging from 7.583 to 10.897 years (Table 2); focusing solely on this metric, it is reasonable to state that, on its own, different anatomical regions perform similarly in age estimation. The same can be said for the metrics of bias, validity, and efficiency. Predictive interval width is perhaps the most distinctive metric for practical applications. Anatomical regions with strong developmental signs, such as the clavicle or the pubis, tend to provide narrower predictive intervals for younger individuals.

Combining traits from different regions provided an improvement over models built on specific anatomic regions. Using 16 traits from standard age-related traits—clavicle, first rib, pubic symphysis, sacroiliac complex (auricular surfaces, S1 body surface, and S1-S2 fusion), resulted in a MAE of 6.609 (5.561–7.598, 95% CI) and reduced the prediction bias considerably when compared to any model built on the same anatomical regions independently (Table 2), and a PIW of 34.245 (12.927–41.087, PIW 95% CI). A model based only on degenerative traits (m = 39) resulted in a MAE of 6.962 (6.084–7.814, 95% CI) and median PIW of 33.732 (28.882–33.122, PIW 95% CI). From our results, multifactorial age estimation models provide improved efficiency, as reflected in narrower predictive intervals (Figures 3–5).

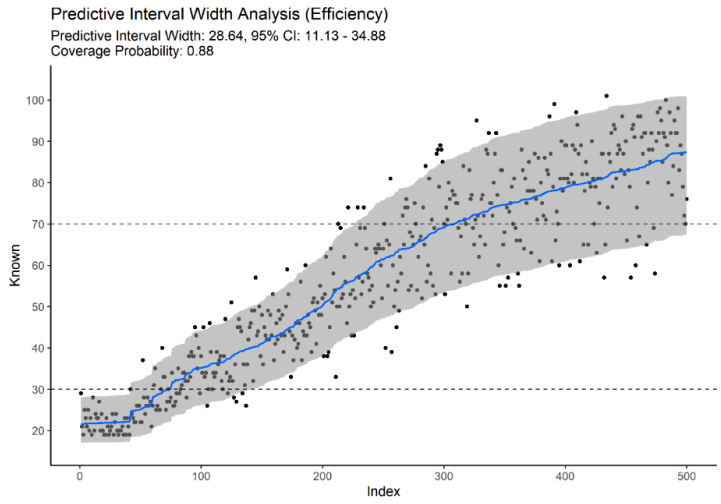


Figure 3. Predictive efficiency of standard age-related traits, $\alpha = 0.1$.

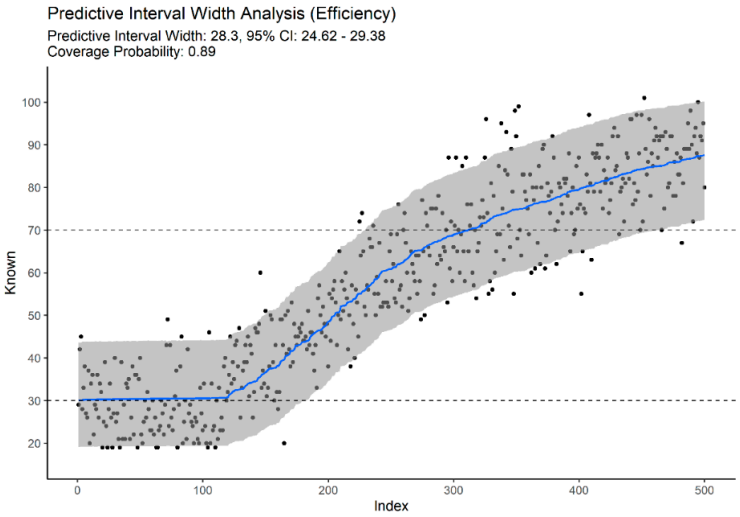


Figure 4. Predictive efficiency of degenerative traits of the axial and appendicular skeleton, $\alpha = 0.1$.

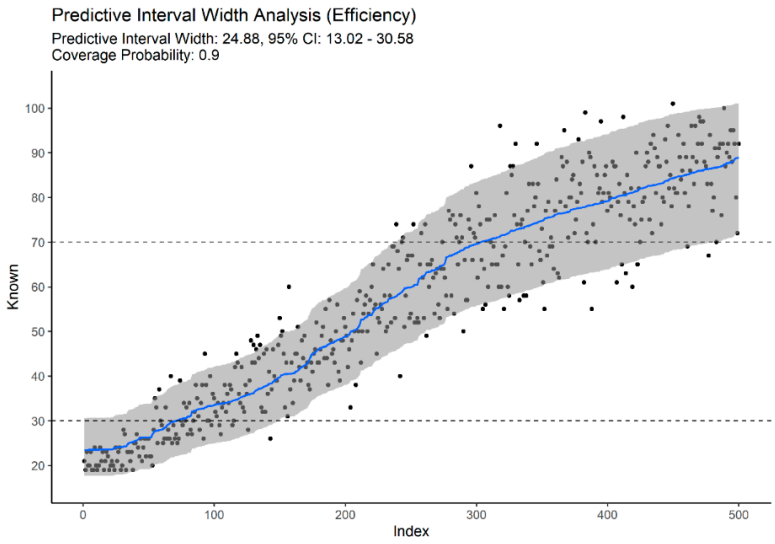


Figure 5. Predictive efficiency of full traits, DRNN-RUM model, $\alpha = 0.1$.

From Figures 3–5, we can also observe that multifactorial models provide accurate and efficient estimates across the entire adult lifespan, solving the problem of open-ended and unspecific age-at-death estimates for the elderly. Figure 4 illustrates the importance of non-standard traits to accurately predict advanced age-at-death. Based solely on degenerative traits of the vertebrae, limb joint, and musculoskeletal attachment sites, we can obtain estimates for the elderly that are comparable to more classical traits (Figure 3) or full-set models (Figure 5). The downside of relying solely on this type on indicator for age-at-death estimation is the wider intervals for young adults with no degenerative traits (95% PI ~18 to 46 years vs. ~18 to 32 if traits with sharp developmental stages are present).

The best performing models in experiment A were those built on the full feature set (m = 64), with a mean absolute error of 5.925 (5.110–6.728, 95% CI), and PIW of 30.010 (15.63–36.081, PIW 95% CI) years. The prediction bias for this model was 0.117 (0.060–0.170, 95% CI), which represents a two-to-six-fold reduction in the prediction bias compared to other models built on specific anatomical regions individually (Table 2). Results from experiment B (Tables 4 and 5) showed that similar results can be obtained using different proportions of traits selected at random.

An important remark to make regarding our results based on the two computational experiments is that analytical LOOCV, implicitly performed during model optimization, showed little to no disparity with the results obtained during the repeats of the Monte Carlo cross-validation procedure (B = 1000 repeats) where 20% of the data was used as a proper test set.

The accuracy of our approach can be visualized in Figure 6, where a scatter plot of known vs. predicted age-at-death is depicted. From this figure, one can infer that the predictions obtained using our approach maintain a similar level of error—dispersion around the identity line (dashed red line)—across the entire adult age span, and slightly more accurate for individuals under 40 years. For individuals over 90 years old at death, there is an observable under-estimation. It is also possible to visualize, Figure 7, that a deep RANN model using multiple traits produces minimally biased estimates.

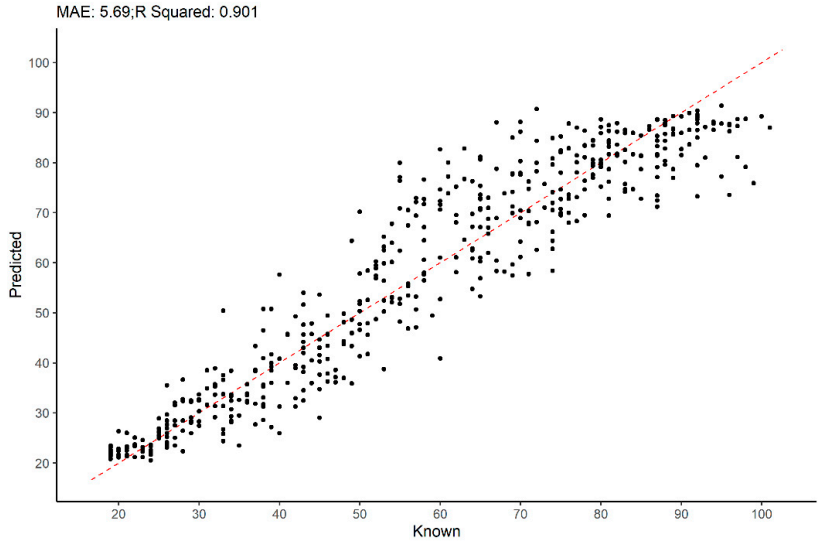


Figure 6. Known vs. predicted age-at-death using a full set of traits (LOOCV, n = 500).

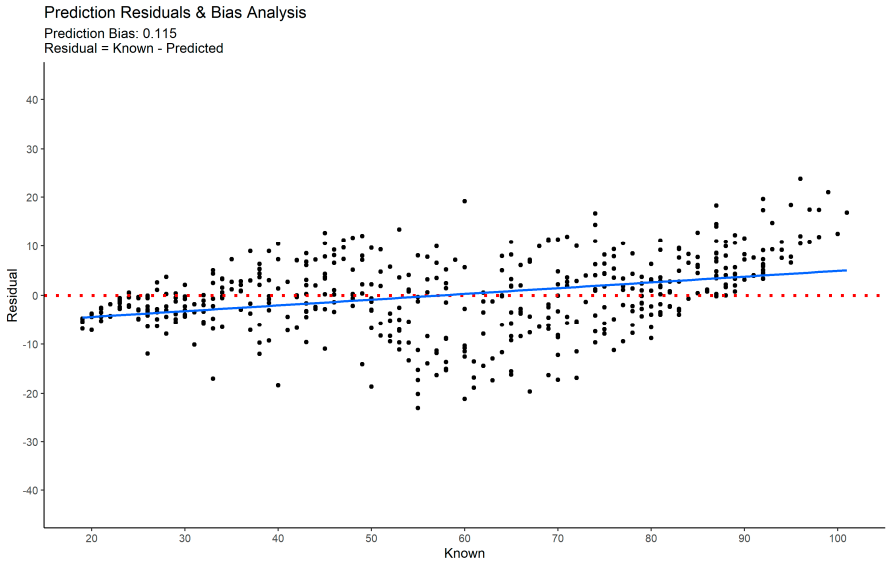


Figure 7. Prediction bias plot for the multifactorial (m = 64) RANN model.

Regarding the validity of the models trained in our computational experiments, results show that the predictive intervals contained the known age-at-death without significant deviation from the nominal level of uncertainty (median of $P(\alpha) \sim 0.95$, with variation between 0.87 and 0.99). Multifactorial models also show a systematical reduction in prediction bias when compared to models based only on a specific anatomical structure.

4. Discussion

The main objective of this work was to investigate the fundamental issue of age-at-death estimation in the forensic analysis of human remains, and propose a new method and its computational analysis from a perspective of multifactorial analysis of the adult skeleton. Several age estimation methods have been previously developed, focusing on specific anatomical structures or regions such as the cranium, the ribs, or the pelvic joints. Nonetheless, it is well known that no single skeletal indicator is capable of producing accurate and efficient age estimates across the entire human age span. Determining how to report age estimates using multiple indicators or traits remains an open issue, with experts resorting to different heuristics that often are not standardized and lack a valid computational or statistical grounding [5]. In the literature, there are techniques that use multiple skeletal indicators for age estimation but are often limited to the cranial sutures and the pelvic joints [20,23,132]. More generic procedures for multifactorial analysis have also been proposed [133,134], but with poor adoption in forensic casework because they require seriation or advanced mathematical knowledge to be put into action.

The current study provides strong support for multifactorial or multi-trait analysis of the skeleton as a way of obtaining accurate and efficient age estimates across the entire span of adulthood. Results from computational experiment A suggest that using each skeletal indicator or anatomical region separately provides limited improvement over existing methods. One striking remark from this experiment was the performance of the models solely based on the axial (vertebrae) and appendicular (limbs) skeleton. In previous studies, these traits have been considered to be only useful for providing a general estimate or limited in value for age prediction [135,136]; nonetheless, our results are consistent with those of more recent publications that assess their predictive utility and urge reconsideration of these traits as valid age-related traits [64,66]. For instance, if these traits all present a Stage 0, one can infer without any computation that the age-at-death of the deceased is between approximately 18 and 46 years (Figure 4, considering $\sigma = 0.1$). Our results also indicate that the inclusion of these traits is pivotal to solve the problem of open-ended age intervals and poor age estimation for the elderly. On their

own, degenerative axial and appendicular traits allow estimation of the age-at-death of the elderly with an improved accuracy and efficiency compared to more standard traits such as the pelvic joints (i.e., pubic symphysis, acetabulum, iliac auricular surface). The neural model based on the full set of traits described in the novel macroscopic age estimation proposed here provided the best performance results in respect to all metrics analyzed. This can be attributed to the fact that having more features allows the deep neural models to operate at their maximum potential regarding what they do best—extracting novel features from existing ones using, in our case, random weights and a non-linearity (ReLU function) as a mechanism to combine multiple traits, which ultimately allows the output layer to operate in a non-linear regime, despite it being, in practice, a regularized linear model. Moreover, the multitude of traits scored also permits the models to encapsulate the intra- and inter-variability of skeletal morphology with greater finesse, which is manifested as more efficient (narrower) predictive intervals that reflect the heteroskedastic nature associated with the senescence process.

Although the main goal of the computational experiment A was to establish a baseline of performance of multifactorial age-at-death estimation compared to more traditional modeling approaches based on specific anatomical blocks or regions, experiment B aimed to assess the performance of neural models for age-at-death estimation in a more realistic setting, where the expert may not be able to use the pre-specified models or the full set of traits due to the availability of skeletal elements or the multitude of factors that make it impossible to score all traits defined in this macroscopic technique. This computational experiment also provides, both directly and indirectly, answers to several questions that may arise regarding the approach and technique used, and proposed in this work from a more pragmatical and casework view: Does the skeleton need to be complete to reap the maximum benefits of this protocol? Which combination of traits works best or is necessary? How practical is the method?

The results demonstrated that the accuracy of the full-set model (m = 64) can be maintained to large degree using smaller random combinations of traits, which ultimately are dictated on a case-by-case basis in a forensic setting. Once again, this can be explained by the capacity of the neural models to extract and combine information from the skeletal traits in an optimal way in terms of prediction. It is important to note here that models based on randomized proportions of traits presented performance metrics superior to most models based on specific anatomical regions, which reinforces our thesis that the multifactorial or multi-trait models are crucial for improving the state-of-art in forensic skeletal age estimation.

Finding an optimal or minimum number of traits is, from a combinatorial and practical point of view, an intractable problem, for which a solution can only be approximated with such a large number of traits (m = 64). However, such a solution would be computational wasteful and of little pragmatic value because, as in the situation of the full trait set, the optimal or minimum trait set can result in a non-applicable model due to the availability of skeletal elements during casework. This is the main reason why, in our study, we opted for a randomized evaluation of smaller traits set. Ultimately, we developed the DRNNAGE software to operationalize the age estimation procedure described in this manuscript, in a manner that is flexible and practical for the expert applying it, bearing in mind that each case will be limited by its own available skeletal traits. DRNNAGE allows the expert to compute the optimal network and associated uncertainty model based only on the traits that the forensic expert can score. Thus, in that regard, the usefulness of the estimates obtained is limited by biology and taphonomy itself, rather than the technical implementation.

From a practitioner perspective, marginal correlation analysis and the performance of the developed models clearly suggest that there is room for improvement in our approach regarding the issue of the traits to be used. For instance, our results suggest that there is little to be gained from including the cranial sutures, which, from a predictive modeling standpoint, resulted in the worst model on its own using our scoring protocol. Similar conclusions were reached by Jooste et al. [137], who also investigated the cranial sutures in the

context of a multifactorial approach. To maximize the potential of the framework proposed in this work, it is important to bear in mind that domain and expert knowledge is of utmost importance; this can also be said of any other machine learning or computationally heavy approach. The practical aspect of this method can be improved if applied with the rationale of the well-known Two-Step Procedure proposed by Baccino et al. [138]. This procedure and heuristic for age-at-death estimation suggests age indicators should be combined logically or hierarchically rather than by brute force (i.e., averaging). In the context of our proposal, this translates into the following: if several traits with sharp metamorphic or developmental stages exhibit Stage 0—i.e., clavicle sternal end, S1-S2 fusion, pubic symphysis components—a neural model is trained using those traits and the other traits are ignored. The same rationale can be applied if the traits that encode a strong degenerative signal, such as the vertebrae and limb traits, are scored with their maximum stage (Stages 1 or 2). In this case, we have demonstrated that age estimation can be accurate and efficient when relying solely on these traits. As a final remark and suggestion to improve age estimation with our method, but also with any other method that employs a multifactorial or multi-trait approach, rather than focusing on an optimal or minimal number of traits to use, one should focus on the representational power of the traits analyzed and, whenever possible, use traits that represent both metamorphic and degenerative aspects of the skeletal development and senescence, as argued by Winburn [88].

The present work provides a solution to the problem of multifactorial age estimation based on the macroscopic analysis of the skeleton. Multifactorial skeletal age estimation is systematically noted as being the most accurate way to achieve an age estimation in adults, but is obtained through a plethora of procedures and heuristics that are often subjective and lack a clearly well-defined statistical or computational rationale [3,5]. As noted by Ritz-Timme et al. [3], a comparison of different methods with regard to their performance based on published data is an exercise that can only be undertaken with severe limitations and caution. The existing methods have been developed on samples of differing sizes, unbalanced age distributions, and different population backgrounds. There is no standardized array of statistical parameters used to assess an age estimation method, and different statistical procedures have been applied. In many cases, there is a lack of detail regarding the procedures used, and often only an incomplete analysis performance is pursued (i.e., focusing only on MAE and point estimate accuracy). In the context of our research subject, these limitations are exacerbated by the fact that, to the best of our knowledge, no other study in the literature has pursued a systematic analysis of adult skeletal age estimation using such a vast and diverse array of morphoscopic traits based on a single reference dataset. Nonetheless, a brief analysis of the most recent and comprehensive validation studies clearly demonstrates that our multifactorial approach offers improved accuracy (MAE < 8 years) in relation to other skeletal age estimation methods [137,139–141]. Independent validation of the method and software tools proposed here on samples from different temporal and biogeographic origins are of utmost importance to ascertain the broader impact and significance in archaeology, forensic anthropology, and medicine.

Artificial intelligence, statistical, and machine learning approaches are now ubiquitous in forensic and biological sciences. Several cases in the literature illustrate the usefulness of such approaches in adult macroscopic age-at-death estimation [13–15,22,24,124]. Although these approaches usually allow for flexible and non-parametric modeling with improved predictive performance, it also results in more opaque or black-box models from a non-expert perspective. These approaches also require proper validation and model selection techniques to avoid overfitting [142]. In this study, we applied a resampling approach to cross-validation based on Monte Carlo cross-validation for fair model assessment, and we also used a robust, analytical, computationally efficient leave-one-out cross-validation strategy to set the regularization parameter of the networks developed in experiments A and B. Randomization rather than optimization of the hidden layers, combined with an efficient C++ implementation of our models, allowed the construction of software that

enables on-the-fly computation and validation (LOOCV) of deep architecture models for any combination of traits with minimal to no technical knowledge on the part of the user.

The problem of interpretability and explainability is a current issue in computational systems using machine learning techniques and constitutes an active topic of research in artificial intelligence [143]. A detailed methodological and implementation analysis will be the focus of a future work, but we briefly describe here how we handle the issue of explainability and interpretability in age-at-death using the neural networks with our software. As previously stated, we can look at the neural network fitted using the techniques described in this manuscript as a regularized linear model operating on the non-linear features extracted by the hidden layers concatenated with the original input (skip layer). We can exploit this property and use the intuitive and additive nature intrinsic to linear models and build a linear surrogate model to explain or interpret any neural network and its predictions.

In DRNNAGE, we regress the cross-validated predictions of the DRNN model on the original input of the network. We decorrelate the input data using the previously described sphering technique and standardize it to zero mean and unit variance. This results in a surrogate model where the intercept or baseline is the average of network estimates, and a new estimate can be "explained" by the sum of the contributions of individual traits to arrive at an approximation of the network estimate (Figure 8).

Linear Approximation

Baseline	Estimate	Approximation
57.336	22.409	20.679

Trait Contribution

Trait	Value	Contribution	Rank
CLV01	0	-19.542	1
RB101	0	-11.425	2
SC01	0	-3.068	3
HM02	0	-1.473	4
HM01	0	-1.149	5

Figure 8. Explanation of an estimate by a linear surrogate model as performed by DRNNAGE software.

Our results suggest that a regression-based framework produces accurate age estimation in adult individuals. Prediction intervals can be estimated with ease and computational efficiency. Bayesian approaches [16,20,23] could have been used for this purpose but they encapsulate a different philosophy to data analysis and are more restrictive in regard to assumptions, parameterization, and computational efficiency compared to the ANN approach we pursued here. Recent contributions suggest that Bayesian approaches do not radically improve age-at-death estimation or outperform regression-based approaches [144,145].

The predictive modeling or function approximation approach pursued in this work is, at the same time, its strongest point and its key limitation. Although neural networks as function approximation machines allowed us to obtain individual accurate age estimates, a predictive modeling strategy—regardless of the underlying algorithm—can only demonstrate that there is an efficient mapping in the form of $y = f^*(x)$. Such a strategy does not explain the underlying biology of the skeletal traits. Fully understanding the biology of the skeletal traits used in age estimation is perhaps the greatest challenge of

this problem, and perhaps the solution for more refined age estimation based solely on the skeletal morphology.

Despite the promising results, the current research did not emerge in a vacuum, nor has it any pretension to be a one-size-fits-all solution to skeletal age estimation, because it was inspired by significant work that was previously developed on this topic, see [16,19,24,35,140].

An important technical and methodological aspect that deserves a detailed analysis in the future is intra- and interobserver error. The results demonstrate the proposed scoring method is highly reproducible. This can be explained by the fact that most traits are encoded in a binary fashion; nonetheless, more data are required from an independent third party that applies the method as described here.

One last aspect that deserves discussion is the dataset employed in this study. The constructed dataset aimed to be uniform and homogeneous in respect to age-at-death and sex. At the moment, it only represents Portuguese nationals over a broad time span; thus, it would be important to expand the dataset to include individuals from other regions, and ascertain possible population and temporal differences in the performance of the proposed method.

5. Conclusions

The work presented here is an important and valuable contribution to the field of age-at-death estimation. Our results clearly demonstrated that a multifactorial approach improves accuracy and precision over single anatomic regions, as established in traditional adult skeletal aging methods. Multifactorial neural models introduce a two-to-six-fold reduction in the mean absolute error and prediction bias compared to standard models. This research also demonstrated that it is possible to produce informative age estimates for the elderly and that nonstandard skeletal traits are pivotal in the later stage of the adult age span. As an age estimation technique developed with forensic casework as its applicational domain, proper validation by other researchers and practitioners is most needed as we are aware that our results, as solid as they are, reflect only in silico performance and cross-validation. This work clearly demonstrated that neural network models offer excellent predictive accuracy. A current issue to be further investigated in future research work is the problem of interpretability and explainability. We briefly alluded to how this problem can be tackled using a global surrogate modeling approach, but other techniques will be investigated in the future so that age-at-death estimation can be approached with computationally accurate and intelligible techniques.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biology11040532/s1. Table S1: Scoring system for suture obliteration. Table S2: List of cranial and palatine suture segment analyzed. Table S3: Scoring system for S1-S2 fusion. Table S4: Scoring system for vertebral body development and degeneration. Table S5: List of traits analyzed in the cervical, lumbar, and sacral vertebrae. Table S6: List of traits used to assess joint and musculoskeletal degeneration of the limbs. Table S7: Generic scoring system for joint degeneration traits. Table S8: Generic scoring system for musculoskeletal degeneration traits. Table S9: Stage 1 specific descriptions for selected joint and musculoskeletal degeneration traits. Table S10: Scoring system for clavicle age-related traits. Table S11: Scoring system for the first rib age-related traits. Table S12: Scoring system for the pubic symphysis age-related traits. Table S13: Scoring system for the sacral auricular age-related traits. Table S14: Scoring system for the iliac auricular age-related traits. Table S15: Scoring system for the acetabular age-related traits.

Author Contributions: Conceptualization, D.N.; Data curation, D.N.; Methodology, D.N.; Software, D.N.; Supervision, E.C. (Ernesto Costa) and E.C. (Eugénia Cunha); Validation, D.N.; Visualization, D.N.; Writing–original draft, D.N., E.C. (Ernesto Costa) and E.C. (Eugénia Cunha); Writing–review and editing, D.N., E.C. (Ernesto Costa) and E.C. (Eugénia Cunha). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fundação para a Ciência e Tecnologia, grant number SFRH/BD/99676/2014.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used for this manuscript preparation is embedded in the software DRNNAGE and it is available from GitHub, see Section 2.6.4. Raw data is available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dirkmaat, D.C.; Cabo, L.L.; Ousley, S.D.; Symes, S.A. New perspectives in forensic anthropology. *Am. J. Phys. Anthropol.* **2008**, *137*, 33–52. [CrossRef] [PubMed]
2. Dirkmaat, D.C.; Cabo, L.L. Embracing the New Paradigm. In *A Companion to Forensic Anthropology*, 1st ed.; Dirkmaat, D.C., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA; pp. 1–40.
3. Ritz-Timme, S.; Cattaneo, C.; Collins, M.; Waite, E.R.; Schütz, H.W.; Kaatsch, H.-J.; Borrman, H.I.M. Age estimation: The state of the art in relation to the specific demands of forensic practise. *Int. J. Legal Med.* **2000**, *113*, 129–136. [CrossRef] [PubMed]
4. Ferrante, L.; Cameriere, R. Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. *Int. J. Legal Med.* **2009**, *123*, 277–283. [CrossRef] [PubMed]
5. Garvin, H.M.; Passalacqua, N.V. Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *J. Forensic Sci.* **2011**, *57*, 427–433. [CrossRef] [PubMed]
6. Franklin, D. Forensic age estimation in human skeletal remains: Current concepts and future directions. *Leg. Med.* **2010**, *12*, 1–7. [CrossRef] [PubMed]
7. Rösing, F.; Graw, M.; Marré, B.; Ritz-Timme, S.; Rothschild, M.; Rötzscher, K.; Schmeling, A.; Schröder, I.; Geserick, G. Recommendations for the forensic diagnosis of sex and age from skeletons. *HOMO J. Comp. Hum. Biol.* **2007**, *58*, 75–89. [CrossRef]
8. Kimmerle, E.H.; Prince, D.A.; Berg, G.E. Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth. *J. Forensic Sci.* **2008**, *53*, 594–600. [CrossRef]
9. Martrille, L.; Ubelaker, D.H.; Cattaneo, C.; Seguret, F.; Tremblay, M.; Baccino, E. Comparison of Four Skeletal Methods for the Estimation of Age at Death on White and Black Adults. *J. Forensic Sci.* **2007**, *52*, 302–307. [CrossRef]
10. Buckberry, J. The (mis)use of adult age estimates in osteology. *Ann. Hum. Biol.* **2015**, *42*, 323–331. [CrossRef]
11. Bocquet-Appel, J.-P.; Masset, C. Farewell to paleodemography. *J. Hum. Evol.* **1982**, *11*, 321–333. [CrossRef]
12. Samworth, R.; Gowland, R. Estimation of adult skeletal age-at-death: Statistical assumptions and applications. *Int. J. Osteoarchaeol.* **2006**, *17*, 174–188. [CrossRef]
13. Kotěrová, A.; Navega, D.; Štepanovský, M.; Buk, Z.; Brůžek, J.; Cunha, E. Age estimation of adult human remains from hip bones using advanced methods. *Forensic Sci. Int.* **2018**, *287*, 163–175. [CrossRef] [PubMed]
14. Navega, D.; Costa, E.; Cunha, E. Lost in the woods: The value of tree ensemble modelling for adult age-at-death estimation from skeletal degeneration. *La Rev. Med. Légale* **2017**, *8*, 181–182. [CrossRef]
15. Navega, D.; Coelho, J.D.; Cunha, E.; Curate, F. DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. *J. Forensic Sci.* **2017**, *63*, 497–503. [CrossRef]
16. Konigsberg, L.W. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann. Hum. Biol.* **2015**, *42*, 368–378. [CrossRef]
17. Lucy, D.; Aykroyd, R.G.; Pollard, A.M. Nonparametric calibration for age estimation. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2002**, *51*, 183–196. [CrossRef]
18. Konigsberg, L.; Frankenberg, S.R. Estimation of age structure in anthropological demography. *Am. J. Phys. Anthropol.* **1992**, *89*, 235–256. [CrossRef]
19. Boldsen, J.L.; Milner, G.R.; Konigsberg, L.W.; Wood, J.W. Transition analysis: A new method for estimating age from the skeletons. In *Paleodemography: Age Distributions from Skeletal Samples*, 1st ed.; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK; pp. 73–106.
20. Milner, G.R.; Boldsen, J.L. Transition analysis: A validation study with known-age modern American skeletons. *Am. J. Phys. Anthropol.* **2012**, *148*, 98–110. [CrossRef]
21. Steadman, D.W.; Adams, B.J.; Konigsberg, L.W. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.* **2006**, *131*, 15–26. [CrossRef]
22. Corsini, M.-M.; Schmitt, A.; Bruzek, J. Aging process variability on the human skeleton: Artificial network as an appropriate tool for age at death assessment. *Forensic Sci. Int.* **2005**, *148*, 163–167. [CrossRef]
23. Martins, R.; Oliveira, P.E.; Schmitt, A. Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. *Forensic Sci. Int.* **2012**, *219*, 287.e1–287.e7. [CrossRef] [PubMed]
24. Buk, Z.; Kordik, P.; Bruzek, J.; Schmitt, A.; Snorek, M. The age at death assessment in a multi-ethnic sample of pelvic bones using nature-inspired data mining methods. *Forensic Sci. Int.* **2012**, *220*, 294.e1–294.e9. [CrossRef] [PubMed]
25. Baccino, E.; Ubelaker, D.H.; Hayek, L.-A.C.; Zerilli, A. Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. *J. Forensic Sci.* **1999**, *44*, 931–936. [CrossRef] [PubMed]

26. Brooks, S.; Suchey, J.M. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* **1990**, *5*, 227–238. [CrossRef]
27. Hanihara, K.; Suzuki, T. Estimation of age from the pubic symphysis by means of multiple regression analysis. *Am. J. Phys. Anthropol.* **1978**, *48*, 233–239. [CrossRef] [PubMed]
28. Gilbert, B.M.; McKern, T.W. A method for aging the femaleOs pubis. *Am. J. Phys. Anthropol.* **1973**, *38*, 31–38. [CrossRef] [PubMed]
29. McKern, T.W.; Stewart, T.D. *Skeletal Age Changes in Young American Males, Analyzed from the Standpoint of Age Identification*; Cambridge University Press: Natick, MA, USA, 1957.
30. Todd, T.W. Age changes in the pubic bone. I. The male white pubis. *Am. J. Phys. Anthropol.* **1920**, *3*, 285–334. [CrossRef]
31. Todd, T.W. Age changes in the pubic bone. *Am. J. Phys. Anthropol.* **1921**, *4*, 1–70. [CrossRef]
32. Katz, D.; Suchey, J.M. Age determination of the male Os pubis. *Am. J. Phys. Anthropol.* **1986**, *69*, 427–435. [CrossRef]
33. Stoyanova, D.; Algee-Hewitt, B.F.; Slice, D.E. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am. J. Phys. Anthropol.* **2015**, *158*, 431–440. [CrossRef]
34. Slice, D.E.; Algee-Hewitt, B.F. Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. *J. Forensic Sci.* **2015**, *60*, 835–843. [CrossRef] [PubMed]
35. Milner, G.R.; Boldsen, J.L. Skeletal Age Estimation: Where We Are and Where We Should Go. In *A Companion to Forensic Anthropology*, 1st ed.; Dirkmaat, D.C., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2017; pp. 224–238.
36. Milner, G.R.; Boldsen, J.L. Estimating Age and Sex from the Skeleton, a Paleopathological Perspective. In *A Companion to Paleopathology*, 1st ed.; Grauer, A.L., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2008; pp. 268–284.
37. Vilas-Boas, D.; Wasterlain, S.; d'Oliveira Coelho, J.; Navega, D.; Gonçalves, D. SPINNE: An app for human vertebral height estimation based on artificial neural networks. *Forensic Sci. Int.* **2019**, *298*, 121–130. [CrossRef] [PubMed]
38. Scott, G.R.; Pilloud, M.A.; Navega, D.; Coelho, J.D.; Cunha, E.; Irish, J.D. rASUDAS: A New Web-Based Application for Estimating Ancestry from Tooth Morphology. *Forensic Anthropol.* **2018**, *1*, 18–31. [CrossRef]
39. Navega, D.; Coelho, C.; Vicente, R.; Ferreira, M.T.; Wasterlain, S.; Cunha, E. AncesTrees: Ancestry estimation with randomized decision trees. *Int. J. Leg. Med.* **2014**, *129*, 1145–1153. [CrossRef]
40. Damas, S.; Wilkinson, C.; Kahana, T.; Veselovskaya, E.; Abramov, A.; Jankauskas, R.; Jayaprakash, P.; Ruiz, E.; Navarro, F.; Huete, M.; et al. Study on the performance of different craniofacial superimposition approaches (II): Best practices proposal. *Forensic Sci. Int.* **2015**, *257*, 504–508. [CrossRef]
41. Mesejo, P.; Martos, R.; Ibáñez; Novo, J.; Ortega, M. A Survey on Artificial Intelligence Techniques for Biomedical Image Analysis in Skeleton-Based Forensic Human Identification. *Appl. Sci.* **2020**, *10*, 4703. [CrossRef]
42. Cunha, E.; Wasterlain, S. The Coimbra identified osteological collections. In *Skeletal Series and Their Socio-Economic Context (Documenta Archaeobiologiae; Bd. 5)*, 1st ed.; Grupe, G., Peters, J., Eds.; Verlag Marie Leidorf GmbH, Rahden/Westf.: Rahden, Germany, 2013; pp. 23–34.
43. Ferreira, M.T.; Vicente, R.; Navega, D.; Gonçalves, D.; Curate, F.; Cunha, E. A new forensic collection housed at the University of Coimbra, Portugal: The 21st century identified skeletal collection. *Forensic Sci. Int.* **2014**, *245*, 202.e1–202.e5. [CrossRef]
44. Ferreira, M.T.; Coelho, C.; Makhoul, C.; Navega, D.; Gonçalves, D.; Cunha, E.; Curate, F. New data about the 21st Century Identified Skeletal Collection (University of Coimbra, Portugal). *Int. J. Leg. Med.* **2020**, *135*, 1087–1094. [CrossRef]
45. Usher, B.M. Reference samples: The first step in linking biology and age in the human skeleton. In *Paleodemography*; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 29–47.
46. Beretta, L.; Santaniello, A. Nearest neighbor imputation algorithms: A critical evaluation. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, 197–208. [CrossRef]
47. Kemkes-Grottenthaler, A. Aging through the ages: Historical perspectives on age indicators methods. In *Paleodemography: Age Distributions from Skeletal Samples*, 1st ed.; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK, 2016; pp. 48–72.
48. Shirley, N.R.; Montes, P.A.R. Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase versus Component-Based Methods. *J. Forensic Sci.* **2014**, *60*, 107–111. [CrossRef]
49. Todd, T.W.; Lyon, D.W. Cranial Suture Closure, Its Progress and Age Relationship. Part I. *Am. J. Phys. Anthropol.* **1925**, *8*, 325–384.
50. Todd, T.W.; Lyon, D.W. Cranial suture closure Part II. *Am. J. Phys. Anthropol.* **1925**, *9*, 23–44. [CrossRef]
51. Meindl, R.S.; Lovejoy, O. Ectocranial suture closure: A revised method for the determination of skeletal age at death and blind tests of its accuracy. *Am. J. Phys. Anthropol.* **1985**, *68*, 57–66. [CrossRef] [PubMed]
52. Perizonius, W.R.K. Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (a.d. 1883–1909). *J. Hum. Evol.* **1984**, *13*, 201–216. [CrossRef]
53. Mann, R.W.; Jantz, R.L.; Bass, W.M.; Willey, P.S. Maxillary Suture Obliteration: A Visual Method for Estimating Skeletal Age. *J. Forensic Sci.* **1991**, *36*, 781–791. [CrossRef]
54. Mann, R.W.; Symes, S.A.; Bass, W.M. Maxillary Suture Obliteration: Aging the Human Skeleton Based on Intact or Fragmentary Maxilla. *J. Forensic Sci.* **1987**, *32*, 148–157. [CrossRef]
55. Acsadi, J.; Nemeskeri, G. *History of Human Life Span and Mortality*; Académiai Kiadó: Budapest, Hungary, 1970.
56. Masset, C. Age estimation on the basis of cranial sutures. In *Age Markers in the Human Skeleton*; İşcan, M.Y., Ed.; Charles C Thomas: Springfield, IL, USA, 1989; pp. 71–103.
57. Ríos, L.; Weisensee, K.; Rissech, C. Sacral fusion as an aid in age estimation. *Forensic Sci. Int.* **2008**, *180*, 111.e1–111.e7. [CrossRef]

58. Belcastro, M.G.; Rastelli, E.; Mariotti, V. Variation of the degree of sacral vertebral body fusion in adulthood in two European modern skeletal collections. *Am. J. Phys. Anthropol.* **2007**, *135*, 149–160. [CrossRef]
59. Passalacqua, N.V. Forensic Age-at-Death Estimation from the Human Sacrum. *J. Forensic Sci.* **2009**, *54*, 255–262. [CrossRef]
60. Snodgrass, J.J. Sex Differences and Aging of the Vertebral Column. *J. Forensic Sci.* **2004**, *49*, 1–6. [CrossRef]
61. Watanabe, S.; Terazawa, K. Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. *Leg. Med.* **2006**, *8*, 156–160. [CrossRef] [PubMed]
62. Albert, M.; Mulhern, D.; Torpey, M.A.; Boone, E. Age Estimation Using Thoracic and First Two Lumbar Vertebral Ring Epiphyseal Union. *J. Forensic Sci.* **2010**, *55*, 287–294. [CrossRef] [PubMed]
63. Alves-Cardoso, F.; Assis, S. Can osteophytes be used as age at death estimators? Testing correlations in skeletonized human remains with known age-at-death. *Forensic Sci. Int.* **2018**, *288*, 59–66. [CrossRef] [PubMed]
64. Milella, M.; Belcastro, M.G.; Mariotti, V.; Nikita, E. Estimation of adult age-at-death from entheseal robusticit.y: A test using an identified Italian skeletal collection. *Am. J. Phys. Anthropol.* **2020**, *173*, 190–199. [CrossRef]
65. Milella, M.; Belcastro, M.G.; Zollikofer, C.P.; Mariotti, V. The effect of age, sex, and physical activity on entheseal morphology in a contemporary Italian skeletal collection. *Am. J. Phys. Anthropol.* **2012**, *148*, 379–388. [CrossRef]
66. Winburn, A.P.; Stock, M.K. Reconsidering osteoarthritis as a skeletal indicator of age at death. *Am. J. Phys. Anthropol.* **2019**, *170*, 459–473. [CrossRef]
67. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. Effects of osteoarthritis on age-at-death estimates from the human pelvis. *Am. J. Phys. Anthropol.* **2018**, *167*, 3–19. [CrossRef]
68. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. Principal component analysis in the evaluation of osteoarthritis. *Am. J. Phys. Anthropol.* **2016**, *162*, 476–490. [CrossRef]
69. Brennaman, A.L.; Love, K.R.; Bethard, J.D.; Pokines, J.T. A Bayesian Approach to Age-at-Death Estimation from Osteoarthritis of the Shoulder in Modern North Americans. *J. Forensic Sci.* **2016**, *62*, 573–584. [CrossRef]
70. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. The relationship of age, activity, and body size on osteoarthritis in weight-bearing skeletal regions. *Int. J. Paleopathol.* **2018**, *22*, 45–53. [CrossRef]
71. Buikstra, J.E.; Ubelaker, D.H. Standards for Data Collection from Human Skeletal Remains. *Fayetteville* **1995**. [CrossRef]
72. Henderson, C.Y.; Mariotti, V.; Pany-Kucera, D.; Villotte, S.; Wilczak, C. Recording Specific Entheseal Changes of Fibrocartilaginous Entheses: Initial Tests Using the Coimbra Method. *Int. J. Osteoarchaeol.* **2012**, *23*, 152–162. [CrossRef]
73. Falys, C.G.; Prangle, D. Estimating age of mature adults from the degeneration of the sternal end of the clavicle. *Am. J. Phys. Anthropol.* **2014**, *156*, 203–214. [CrossRef]
74. Langley-Shirley, N.; Jantz, R.L. A Bayesian Approach to Age Estimation in Modern Americans from the Clavicle. *J. Forensic Sci.* **2010**, *55*, 571–583. [CrossRef] [PubMed]
75. Owings, W.P.A.; Myers, S.J.; Webb, P.A.O.; Suchey, J.M. Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *Am. J. Phys. Anthropol.* **1985**, *68*, 457–466. [CrossRef]
76. Cardoso, H.F.V. V Age estimation of adolescent and young adult male and female skeletons II, epiphyseal union at the upper limb and scapular girdle in a modern Portuguese skeletal sample. *Am. J. Phys. Anthropol.* **2008**, *137*, 97–105. [CrossRef]
77. Can, M.Y.; Loth, S.R. Determination of age from the sternal rib in white males: A test of the phase method. *J. Forensic Sci.* **1986**, *31*, 122–132.
78. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Age Estimation from the Rib by Phase Analysis: White Males. *J. Forensic Sci.* **1984**, *29*, 1094–1104. [CrossRef]
79. Can, M.Y.; Loth, S.R. Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *Am. J. Phys. Anthropol.* **1984**, *65*, 147–156.
80. Işcan, M.Y.; Loth, S.R.; Wright, R.K. Racial Variation in the Sternal Extremity of the Rib and Its Effect on Age Determination. *J. Forensic Sci.* **1987**, *32*, 452–466. [CrossRef]
81. Kunos, C.A.; Simpson, S.W.; Russell, K.F. Hershkovitz I First rib metamorphosis: Its possible utility for human age-at-death estimation. *Am. J. Phys. Anthropol.* **1999**, *110*, 303–323. [CrossRef]
82. DiGangi, E.A.; Bethard, J.D.; Kimmerle, E.H.; Konigsberg, L.W. A new method for estimating age-at-death from the first rib. *Am. J. Phys. Anthropol.* **2009**, *138*, 164–176. [CrossRef] [PubMed]
83. Sashin, D.A. Critical analysis pf the anatomy and the pathologic changes of the sacro-iliac joints. *J. Bone Jt. Surg.* **1992**, *12*, 891–910.
84. Schunke, G.B. The anatomy and development of the sacro-iliac joint in man. *Anat. Rec.* **1938**, *72*, 313–331. [CrossRef]
85. Lovejoy, C.O.; Meindl, R.S.; Pryzbeck, T.R.; Mensforth, R.P. Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *Am. J. Phys. Anthropol.* **1985**, *68*, 15–28. [CrossRef]
86. Buckberry, J.; Chamberlain, A. Age estimation from the auricular surface of the ilium: A revised method. *Am. J. Phys. Anthropol.* **2002**, *119*, 231–239. [CrossRef]
87. San-Millán, M.; Rissech, C.; Turbón, D. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *Int. J. Leg. Med.* **2016**, *131*, 501–525. [CrossRef]
88. Winburn, A.P. Validation of the Acetabulum as a Skeletal Indicator of Age at Death in Modern European-Americans. *J. Forensic Sci.* **2018**, *64*, 989–1003. [CrossRef]
89. Mays, S. A Test of a Recently Devised Method of Estimating Skeletal Age at Death using Features of the Adult Acetabulum. *J. Forensic Sci.* **2013**, *59*, 184–187. [CrossRef]

90. Calce, S.E. A new method to estimate adult age-at-death using the acetabulum. *Am. J. Phys. Anthropol.* **2012**, *148*, 11–23. [CrossRef]
91. Rissech, C.; Estabrook, G.F.; Cunha, E.; Malgosa, A. Using the Acetabulum to Estimate Age at Death of Adult Males. *J. Forensic Sci.* **2006**, *51*, 213–229. [CrossRef] [PubMed]
92. Rissech, C.; Estabrook, G.F.; Cunha, E.; Malgosa, A.; Badalló, M.D.C.R. Estimation of Age-at-Death for Adult Males Using the Acetabulum, Applied to Four Western European Populations. *J. Forensic Sci.* **2007**, *52*, 774–778. [CrossRef] [PubMed]
93. Rougé-Maillart, C.; Vielle, B.; Jousset, N.; Chappard, D.; Telmon, N.; Cunha, E. Development of a method to estimate skeletal age at death in adults using the acetabulum and the auricular surface on a Portuguese population. *Forensic Sci. Int.* **2009**, *188*, 91–95. [CrossRef] [PubMed]
94. Rougé-Maillart, C.; Jousset, N.; Vielle, B.; Gaudin, A.; Telmon, N. Contribution of the study of acetabulum for the estimation of adult subjects. *Forensic Sci. Int.* **2007**, *171*, 103–110. [CrossRef] [PubMed]
95. San-Millán, M.; Rissech, C.; Turbón, D. Application of the recent SanMillán–Rissech acetabular adult aging method in a North American sample. *Int. J. Leg. Med.* **2019**, *133*, 909–920. [CrossRef]
96. Kendall, M.G.; Smith, B.B. The Problem of m Rankings. *Ann. Math. Stat.* **1939**, *10*, 275–287. [CrossRef]
97. Zuber, V.; Strimmer, K. High-Dimensional Regression and Variable Selection Using CAR Scores. *Stat. Appl. Genet. Mol. Biol.* **2011**, *10*. [CrossRef]
98. Kessy, A.; Lewin, A.; Strimmer, K. Optimal Whitening and Decorrelation. *Am. Stat.* **2018**, *72*, 309–314. [CrossRef]
99. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2015.
100. Scardapane, S.; Wang, D. Randomness in neural networks: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1200. [CrossRef]
101. Gallicchio, C.; Scardapane, S. Deep Randomized Neural Networks. *Stud. Comput. Intell.* **2020**, *896*, 43–68. [CrossRef]
102. Schmidt, W.F.; Kraaijveld, M.A.; Duin, R.P.W. Feedforward neural networks with random weights. In Proceedings of the 11th IAPR International Conference on Pattern Recognition. Volume II. Conference B: Pattern Recognition Methodology and Systems, The Hague, The Netherlands, 30 August 1992; IEEE Comput. Soc. Press: Piscataway, NJ, USA, 1992; pp. 1–4.
103. Pao, Y.-H.; Takefuji, Y. Functional-link net computing: Theory, system architecture, and functionalities. *Computer (Long Beach Ca.)* **1992**, *25*, 76–79. [CrossRef]
104. Broomhead, D.S.; Lowe, D. Multivariable functional interpolation and adaptive networks. *Complex Syst.* **1988**, *2*, 321–355.
105. Pao, Y.-H.; Park, G.-H.; Sobajic, D.J. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing* **1994**, *6*, 163–180. [CrossRef]
106. Igelnik, B.; Pao, Y.-H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans. Neural Netw.* **1995**, *6*, 1320–1329. [CrossRef] [PubMed]
107. Huang, G.-B. What are Extreme Learning Machines? Filling the Gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle. *Cogn. Comput.* **2015**, *7*, 263–278. [CrossRef]
108. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
109. Huang, G.; Huang, G.; Bin Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [CrossRef]
110. Huang, G.-B. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. *Cogn. Comput.* **2014**, *6*, 376–390. [CrossRef]
111. Wang, L.P.; Wan, C.R. Comments on "The Extreme Learning Machine. *IEEE Trans. Neural Netw.* **2008**, *19*, 1494–1495. [CrossRef]
112. Shao, Z.; Er, M.J. Efficient Leave-One-Out Cross-Validation-based Regularized Extreme Learning Machine. *Neurocomputing* **2016**, *194*, 260–270. [CrossRef]
113. Wang, D.; Wang, P.; Shi, J. A fast and efficient conformal regressor with regularized extreme learning machine. *Neurocomputing* **2018**, *304*, 1–11. [CrossRef]
114. Tissera, M.D.; McDonnell, M. Deep extreme learning machines: Supervised autoencoding architecture for classification. *Neurocomputing* **2016**, *174*, 42–49. [CrossRef]
115. Tang, J.; Deng, C.; Huang, G.-B. Extreme Learning Machine for Multilayer Perceptron. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 809–821. [CrossRef] [PubMed]
116. Zhou, H.; Huang, G.-B.; Lin, Z.; Wang, H.; Soh, Y.C. Stacked Extreme Learning Machines. *IEEE Trans. Cybern.* **2014**, *45*, 2013–2025. [CrossRef]
117. Yu, W.; Zhuang, F.; He, Q.; Shi, Z. Learning deep representations via extreme learning machines. *Neurocomputing* **2015**, *149*, 308–315. [CrossRef]
118. Shi, Q.; Katuwal, R.; Suganthan, P.; Tanveer, M. Random vector functional link neural network based ensemble deep learning. *Pattern Recognit.* **2021**, *117*, 107978. [CrossRef]
119. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 513–529. [CrossRef]
120. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]
121. Bartlett, P.L. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **1998**, *44*, 525–536. [CrossRef]

122. Bartlett, P.L. For valid generalization, the size of the weights is more important than the size of the network. *Adv. Neural Inf. Processing Syst. 9 (NIPS 1996)* **1996**, *23*, 35–39.
123. Allen, D.M. The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. *Technometrics* **1974**, *16*, 125. [CrossRef]
124. Navega, D.; Cunha, E. Extreme learning machine neural networks for adult skeletal age-at-death estimation. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, A., Cattaneo, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 209–225.
125. Milborrow, S. Variance Models in Earth. Available online: http://www.milbo.org/doc/earth-varmod.pdf (accessed on 18 March 2022).
126. Geary, R.C. The Ratio of the Mean Deviation to the Standard Deviation as a Test of Normality. *Biometrika* **1935**, *27*, 310. [CrossRef]
127. Konigsberg, L.W.; Herrmann, N.P.; Wescott, D.J.; Kimmerle, E.H. Estimation and Evidence in Forensic Anthropology: Age-at-Death. *J. Forensic Sci.* **2008**, *53*, 541–557. [CrossRef] [PubMed]
128. Müller, H.-G.; Love, B.; Hoppa, R.D. Semiparametric method for estimating paleodemographic profiles from age indicator data. *Am. J. Phys. Anthropol.* **2001**, *117*, 1–14. [CrossRef] [PubMed]
129. Fieuws, S.; Willems, G.; Larsen-Tangmose, S.; Lynnerup, N.; Boldsen, J.; Thevissen, P. Obtaining appropriate interval estimates for age when multiple indicators are used: Evaluation of an ad-hoc procedure. *Int. J. Leg. Med.* **2015**, *130*, 489–499. [CrossRef]
130. Papadopoulos, H.; Haralambous, H. Reliable prediction intervals with regression neural networks. *Neural Netw.* **2011**, *24*, 842–851. [CrossRef]
131. Lappas, G. Estimating the Size of Neural Networks from the number of available training data. *Neural Netw.* **2007**, 68–77. [CrossRef]
132. Rougé-Maillart, C.; Telmon, N.; Rissech, C.; Malgosa, A.; Rougé, D. The determination of male adult age at death by central and posterior coxal analysis–a preliminary study. *J. Forensic Sci.* **2004**, *49*, 1–7. [CrossRef]
133. Lovejoy, C.O.; Meindl, R.S.; Mensforth, R.P.; Barton, T.J. Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. *Am. J. Phys. Anthropol.* **1985**, *68*, 1–14. [CrossRef]
134. Anderson, M.F.; Anderson, D.T.; Wescott, D.J. Estimation of adult skeletal age-at-death using the Sugeno fuzzy integral. *Am. J. Phys. Anthropol.* **2009**, *142*, 30–41. [CrossRef]
135. Listi, G.A. The Use of Entheseal Changes in the Femur and Os Coxa for Age Assessment. *J. Forensic Sci.* **2015**, *61*, 12–18. [CrossRef] [PubMed]
136. Listi, G.A.; Manhein, M.H. The Use of Vertebral Osteoarthritis and Osteophytosis in Age Estimation. *J. Forensic Sci.* **2012**, *57*, 1537–1540. [CrossRef] [PubMed]
137. Jooste, N.; L'Abbé, E.N.; Pretorius, S.; Steyn, M. Validation of transition analysis as a method of adult age estimation in a modern South African sample. *Forensic Sci. Int.* **2016**, *266*, 580.e1–580.e7. [CrossRef] [PubMed]
138. Baccino, E.; Sinfield, L.; Colomb, S.; Baum, T.P.; Martrille, L. Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. *Forensic Sci. Int.* **2014**, *244*, 247–251. [CrossRef]
139. Miranker, M. A Comparison of Different Age Estimation Methods of the Adult Pelvis. *J. Forensic Sci.* **2016**, *61*, 1173–1179. [CrossRef]
140. Merritt, C.E. Inaccuracy and bias in adult skeletal age estimation: Assessing the reliability of eight methods on individuals of varying body sizes. *Forensic Sci. Int.* **2017**, *275*, 315.e1–315.e11. [CrossRef]
141. Hagelthorn, C.L.; Alblas, A.; Greyling, L. The accuracy of the Transition Analysis of aging on a heterogenic South African population. *Forensic Sci. Int.* **2019**, *297*, 370.e1–370.e5. [CrossRef]
142. Valsecchi, A.; Olivares, J.I.; Mesejo, P. Age estimation in forensic anthropology: Methodological considerations about the validation studies of prediction models. *Int. J. Leg. Med.* **2019**, *133*, 1915–1924. [CrossRef]
143. Burkart, N.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. [CrossRef]
144. Jooste, N.; Pretorius, S.; Steyn, M. Performance of three mathematical models for estimating age-at-death from multiple indicators of the adult skeleton. *Int. J. Leg. Med.* **2021**, 1–13. [CrossRef]
145. Nikita, E.; Nikitas, P. Skeletal age-at-death estimation: Bayesian versus regression methods. *Forensic Sci. Int.* **2019**, *297*, 56–64. [CrossRef] [PubMed]

biology

MDPI

Article

How Do Drugs Affect the Skeleton? Implications for Forensic Anthropology

Nicholas Márquez-Grant [1,*], Elisa Baldini [1], Victoria Jeynes [2], Lucie Biehler-Gomez [3], Layla Aoukhiyad [4], Nicholas V. Passalacqua [5], Gaia Giordano [3], Domenico Di Candia [3] and Cristina Cattaneo [3]

1 Cranfield Forensic Institute, Cranfield University, Bedford MK43 0AL, UK; elisa.baldini.uk@gmail.com
2 Urgent Primary Care, Cwm Taf Morgannwg University Health Board, Cardiff CF14 4XW, UK; v.jeynes@outlook.com
3 Department of Biomedical Sciences for Health, Università degli Studi di Milano, 20122 Milano, Italy; lucie.biehler@unimi.it (L.B.-G.); gaia.giordano1@gmail.com (G.G.); domenico.dicandia@unimi.it (D.D.C.); cristina.cattaneo@unimi.it (C.C.)
4 Pharmacist, Servei de Salut de les Illes Balears, 07003 Palma, Spain; lilitfar@yahoo.es
5 Anthropology and Sociology Department, Western Carolina University, Cullowhee, NC 28723, USA; passala5@gmail.com
* Correspondence: n.marquezgrant@cranfield.ac.uk

Simple Summary: Forensic anthropologists analyze human remains to assist in the identification of the deceased, predominantly by assessing age-at-death, sex, stature, ancestry and any unique identifying features. Whilst methods have been established to create this biological profile of the skeleton, these may be influenced by a number of factors. This paper, for the first time, provides an overview from a reading of the clinical and pharmacological literature to explore whether the intake of drugs can affect the skeleton and whether these may have implications for forensic anthropology casework. In effect, drugs such as tobacco, heroin, and prescription medications can alter bone mineral density, can increase the risk of fractures, destroy bone and changes to the dentition. By considering how drugs can affect the skeleton, forensic anthropologists can be aware of this when attempting to identify the deceased.

Abstract: Forensic anthropologists rely on a number of parameters when analyzing human skeletal remains to assist in the identification of the deceased, predominantly age-at-death, sex, stature, ancestry or population affinity, and any unique identifying features. During the examination of human remains, it is important to be aware that the skeletal features considered when applying anthropological methods may be influenced and modified by a number of factors, and particular to this article, prescription drugs (including medical and non-medical use) and other commonly used drugs. In view of this, this paper aims to review the medical, clinical and pharmacological literature to enable an assessment of those drug groups that as side effects have the potential to have an adverse effect on the skeleton, and explore whether or not they can influence the estimation of age-at-death, sex and other indicators of the biological profile. Moreover, it may be that the observation of certain alterations or inconsistencies in the skeleton may relate to the use of drugs or medication, and this in turn may help narrow down the list of missing persons to which a set of human remains could belong. The information gathered from the clinical and medical literature has been extracted with a forensic anthropological perspective and provides an awareness on how several drugs, such as opioids, cocaine, corticosteroids, non-steroidal anti-inflammatory drugs, alcohol, tobacco and others have notable effects on bone. Through different mechanisms, drugs can alter bone mineral density, causing osteopenia, osteoporosis, increase the risk of fractures, osteonecrosis, and oral changes. Not much has been written on the influence of drugs on the skeleton from the forensic anthropological practitioner perspective; and this review, in spite of its limitations and the requirement of further research, aims to investigate the current knowledge of the possible effects of both prescription and recreational drugs on bones, contributing to providing a better awareness in forensic anthropological practice and assisting in the identification process of the deceased.

Citation: Márquez-Grant, N.; Baldini, E.; Jeynes, V.; Biehler-Gomez, L.; Aoukhiyad, L.; Passalacqua, N.V.; Giordano, G.; Di Candia, D.; Cattaneo, C. How Do Drugs Affect the Skeleton? Implications for Forensic Anthropology. *Biology* **2022**, *11*, 524. https://doi.org/10.3390/biology11040524

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 3 February 2022
Accepted: 18 March 2022
Published: 29 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: forensic anthropology; medication; drugs of abuse; biological profile; human identification

1. Introduction

Amongst the requests forensic anthropologists undertake, one major role is to assist in the identification of the deceased through primarily the analysis of human skeletal remains [1–4]. In this regard, during the post-mortem examination of the remains, the anthropologist may be asked to provide information on the biological profile of the individual; this can include the estimation of age-at-death, sex, stature, ancestry (or population affinity), and identifying any unique features [5]. Age-at-death estimation may involve the assessment of skeletal maturation, dental development, and morphological changes in areas such as the pubic symphysis, the rib end, and the auricular surface of the ilium [6,7]. Biological sex estimation may involve an analysis of the pelvic bones, the skull, possibly complemented by metric data [8]. Stature will be estimated by applying bone measurements to an equation [9]; whilst ancestry may be estimated using morphoscopic or metric analyses [10–12]. The skeleton will also be examined for any identifying features such as non-metric traits, evidence of surgery and pathological conditions, that may assist in narrowing down the list of missing persons whose remains are being analyzed [3,5,13].

However, it is important to remember that skeletal indicators considered for the reconstruction of the biological profile are influenced by a number of factors including age, sex, disease, genetics, lifestyle, diet, and pertinent here, possibly the use of prescription drugs (medical and non-medical) and other commonly used drugs, such as those drugs of abuse. Indeed, the medical literature describes how various drugs can affect the skeleton [14] and thus modify characteristic bone quality, appearance, shape and size of skeletal areas [15], which are used for the reconstruction of the biological profile.

The United Nation Office for Drug and Crime estimates that about 275 million people worldwide made use of drugs at least once in 2019, a number that has been increasing by the millions in recent years [16]. Moreover, according to the World Health Organization (WHO), drug use led to approximately 450,000 deaths in 2015 [17]. These figures, added to the number of people that regularly take (prescribed) drugs for medical reasons, show the scale of the phenomenon and in turn the importance of considering the impact of drugs on the skeleton during forensic anthropological casework.

This theme has not been thoroughly investigated in the context of skeletal analysis in forensic anthropology. To date, published literature in this area has so far explored only a minimal part of these effects. For instance, the investigation of particular bone manifestations of cocaine abuse trough CT scans [18]; discussing how homeostasis can change due to alcohol and drug use, affecting the ability to accurately assess estimation of age-at-death [19,20]; or experimental approaches with human analogues on opioids [21]. The presence of drugs in bones has been studied mainly in skeletal toxicology, where the substance is detected analytically [22–27], but very little has been done macroscopically with imaging or by direct examination of the bones.

The main aim of this paper, therefore, is to present and discuss the potential skeletal effects of different medications and drugs based on a review of the literature. This has two advantages: (1) to consider these possible effects when assessing the biological profile through the estimation of age-at-death, sex, stature, etc. from the skeleton; (2) explore whether any changes to the skeleton may be specific to a particular drug or class of drugs, which may then in turn assist with identification, in particular if the medical history of the deceased is available. Although this review is not exhaustive, the final overall aim is to also provide an awareness for the forensic anthropological practitioner, and highlight the importance of further research on this topic.

2. Materials and Methods

To achieve the aims of this paper, medical, clinical, pharmacological and forensic anthropological literature was researched in several scientific databases; and scientific journals and medical books were accessed. The analysis of the literature was divided into two steps: first, the general relationship between drugs and bone health was investigated; second, specific research was carried out on the different drugs that may have bone involvement as side effects.

The literature search was performed between November 2019 and October 2021 and built from a previous MSc thesis [28], using the keywords "bone/s", "drug/s", "medication/s" on several databases including PubMed (Medline), Scopus, Science Direct and Web of Science, as well as Google. Once specific drugs were identified, a more directed research was run using their names to further investigate their effects on bones and a number of drug databases were consulted including Vademecum (www.vademecum.es, accessed 28 October 2021), the Spanish Agency for Medicine (CIMA https://cima.aemps.es/cima/publico/home.html#quees, accessed 28 October 2021), Substance Abuse and Mental Health Administration (SAMHA https://www.samhsa.gov/, accessed 28 October 2021), Alcohol and Drug Foundation (ADF https://adf.org.au/drug-facts/#list, accessed 28 October 2021), National Cancer Institute (https://seer.cancer.gov/seertools/seerrx/, accessed 28 October 2021), UK Government website (https://www.gov.uk/guidance/find-product-information-abut-medicines, accessed 28 October 2021), and the US Food and Drug Administration (FDA https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm, accessed 28 October 2021). Moreover, the Prescription Drugs and Over-the-Counter (OTC) Drugs identified, the official product label was reviewed to check whether the suspected adverse reaction was consistent with those described in the product label (The European Summary of Product Characteristics (SmPC) and the United States Prescribing information (USPI). No restriction regarding the date of publication was applied. Teeth and oral health were examined briefly as this is the remit of the forensic odontologist, rather than the forensic anthropologist.

The results were summarized and organized in two tables by class and type of drug, showing their reported effect on bones and if any, the area of the skeleton most commonly involved. In addition, it was reported whether they could potentially affect sex and age estimation or any other biological profile parameter.

3. Results

The information collected from the literature shows that commonly used drugs (with the potential for misuse or addiction such as prescription opioid, tobacco and alcohol), prescription drugs and even over-the-counter drugs, if taken long term and/or in high doses, have the potential to cause numerous health issues, including bone modifications at different levels [14].

The most commonly used drugs (with the potential for misuse and addition), defined as psychoactive drugs, can be categorised as stimulants, narcotics (opioids), depressants, hallucinogens and marijuana (cannabis) [29]. As will be seen in later sections of this paper, among stimulants, the principal drugs that can have a detrimental effect on the skeleton are cocaine, amphetamines, and nicotine (i.e., the main component of tobacco). Opioids include morphine and its derivatives, methadone and heroin. Alcohol and others (such as benzodiazepines and barbiturates) are depressant drugs with proved side effects on bones, while amongst hallucinogens, ecstasy can also be associated to bone disease. Opioids can be prescription medications, and along with some over-the-counter medications (i.e., nonsteroidal anti-inflammatory drugs and paracetamol), can lead to addiction and are commonly abused. However, side effects which affect the skeleton can also occur by taking controlled doses of other prescribed drugs that usually do not cause addiction but are extensively used in clinical medicine. These medications include corticosteroids, antiresorptive drugs, gonadotropin releasing hormones (GnRH) agonists, gastric acid suppressants or proton pump inhibitors, thyroid hormones and antiretroviral, antidepressant, antipsychotic,

antiepileptic, antidiabetic, and antithrombotic drugs. These are included in more detail in the following section.

3.1. Effects of Drugs on Bone

This section includes the drug classes that, as a result of the research undertaken for this paper, can have an adverse effect on bone. This paper avoids brand names or trademarks and mainly provides classifications that are either generic or according to effect (therapeutic classification), chemical components or mechanisms (pharmacological classification). Brief definitions are provided, alongside a brief overview of their use and how they can affect the skeleton. For each drug, and whenever applicable, macroscopic bone lesions are described as well as their potential effect on the process of age-at-death and sex estimation in forensic anthropology practice. This review is not extensive, at least in its bibliography, but it provides an insight into how medication and drugs of abuse can modify the skeleton, which is an important consideration for forensic anthropologists. A small mention to dental disease and oral pathology, as well as cartilage, is included at the end. Limitations and interpretations are discussed later.

3.1.1. Cocaine

Cocaine is an alkaloid derived from the leaves of the *Erythroxylum coca* plant. It is currently used as an intraoperative local anaesthetic and vasoconstrictor, but it also represents one of the most common drugs of abuse [30]. Recreational cocaine is often contaminated by various additive compounds, such as levamisole, which can be directly responsible for the effects of the drug and/or its local and systemic complications, or act as a contributing factor [31,32]. Cocaine can be administered through intravenous injection, nasal insufflation (the most common), inhalation (smoking), direct application on mucous membranes or chewed and rubbed on the gum. The way cocaine is administered will influence the effect of the drug on bones [30]. In fact, the intranasal use (insufflation) is responsible for one of the most important effects of cocaine on bones, the cocaine-induced midline destructive lesion (CIMDL), characterized by the destruction of the nasal septum, lateral nasal walls and/or hard palate [33–35]. Rubin [18] defined this condition as any significant bone damage of the midfacial region clearly caused by the use of cocaine and identifiable in human skeletal remains. Its pathogenesis is mainly related to the vasoconstrictive effect of cocaine, leading to ischemic necrosis, combined with the chemical irritation of adulterants, direct trauma from the use of paraphernalia and possible superinfection [34]. Thus, after repetitive and frequent snorting, the blood vessels of the nasal mucosa become atrophic and irritated, resulting in localised ischemia and ultimately in necrosis, erosion and destruction of the osteocartilaginous tissue. Septal perforation tends to be observed first, and the lesion then progresses and involves the nasal lateral walls with saddle-nose or alar deformities, the hard palate with oro-nasal fistulas, and even the maxillary sinuses and orbital walls due to chronic inflammation and infection of the sinuses [36–38]. Rubin [18] considers how forensic anthropologists should consider someone as a cocaine abuser where there is lack of new bone formation to repair the lytic lesions. These destructive lesions are primarily located in the vomer, in the palate (palatine bones) and inferior nasal conchae; with other bones affected being the ethmoid, maxillary sinuses, sphenoid and orbit [18]. One clinical case showed also an extension of CIMDL into the neck area, especially with some destruction and instability of the atlanto-axial joint [39].

3.1.2. Opioids

These are naturally found in the opium poppy and can be prescription medications often referred to as painkillers, although are often used non-medically or recreationally. Their use is widespread, and data has shown that it has been taken illegally since adolescence [40]. Three most commonly used opioids are covered here: morphine, methadone, and heroin.

The use of morphine to manage chronic pain is widespread. However, as it would appear that it inhibits osteoblastic activity and certain hormones such as gonadotropin-releasing hormone (GnRH) [41,42], it has been shown that opioids can induce osteoporosis and thus increase osteoporosis risk fracture [43]. This reduction in bone density and thus leading to osteoporosis has been demonstrated in some human and non-human experimental studies [44], although other factors, leading to this lower bone mass density, need to be considered [45]. The risk of fracture in morphine users also increases, especially in common osteoporotic fractures such as those found at the hip, spine, and forearm; a risk increased by loss of postural balance and falls due to side effects of the drug [46]. This, in turn, although not with all opioids, leads negatively to bone healing, and bone non-union may result [47]. Moreover, as it affects cell proliferation and apoptosis [48], experimental studies on rats have shown that morphine in mothers have effects on the primary and secondary ossification and longitudinal growth of their offspring [48,49].

With regard to methadone, Kim and colleagues [50] investigated the low bone mineral density (BMD) in patients taking part in a methadone maintenance program in Boston. Using dual energy X-ray absorptiometry (DXA) combined with surveys and medical records, the study found that BMD of 83% of the study sample were below normal, with 35% of those within the osteoporosis range, and 48% of those in the osteopenia range. This in turn, resulted in a higher fracture risk for those who were taken methadone [51]. Similar studies have been undertaken on male and female subjects yielding different results, with more significant bone loss in the former than in the latter [52,53]. This association may be related to the effect opioids have on bone metabolism, in particular inhibiting osteoblastic (bone formation) activity [54].

Heroin is made by adding two acetyl groups to the molecule morphine. As heroin can alter several body functions, chronic abusers present with altered bone metabolism and reduced trabecular bone mass, which according to Pedrazzoni et al. [55] is attributable partly to hypogonadism. Wilczek, H., and Stěpán, J. [56] investigated the effects of prolonged use of heroin and noted, focusing on the femoral neck and forearm, that it is associated with accelerated bone turnover, resulting in osteopenia. However, after one year of treatment with methadone, bone turnover rate was restored. In addition, a Spanish study noted the presence of septic arthritis in heroin users, affecting especially the sacroiliac, costoclavicular, hip and shoulder joints [57]. In fact, intravenous drug injection in heroin addicts has been associated with osteomyelitis. In a study by Allison et al. [58], out of 215 patients injecting drugs, 59% had osteomyelitis and 25% septic arthritis. In fact, septic arthritis at the pubic symphysis has been found to have intravenous drug injection as a risk factor [59]. Similar associations with osteomyelitis have been found in other studies in the last decades where joint disease and infectious skeletal lesions have been present, usually in the limbs and sites where the injections have taken place [60]. A number of cases since the 1980s have also reported cervical osteomyelitis in intravenous drug use [61,62].

There are also other drugs in this group, such as Desomorphine, a synthesized opioid from codeine which has been associated with skeletal infections at the site of skin ulcers due to injection, followed by necrosis and gangrene in some cases, and amputation [63,64]. Due to the toxic substances in the manufacturing process of this highly addictive drug, as well as the injectable equipment and hygiene, the risk of infection is much larger and more severe than that of any other drug with the same administration [64]. Some of these drugs have also shown to cause necrosis of the mandible and maxilla [65,66].

3.1.3. Amphetamines

As stimulants, they speed up the transmission between the brain and different parts of the body. There are different types of amphetamines, some being prescribed to treat disorders such as attention deficit hyperactivity disorder (ADHD) and other conditions (https://www.dea.gov/sites/default/files/2020-06/Amphetamines-2020_0.pdf (accessed 29 October 2021)). The most potent form is methamphetamine (METH). The main route of administration is orally, but can also be injected intravenously, or taken by insufflation,

inhalation and suppository. Amphetamines decrease bone mass and strength due to the drug effect on the central nervous system, closely linked to bone metabolism and affecting bone turnover [67]. A strong correlation has been found in the literature between methamphetamine users and lower bone density and osteoporosis [51]. For example, Katsuragawa [68] found a decrease in bone mass and integrity in the calcaneus of drug users. In addition, Mosti and colleagues [69] examined loss of bone density by assessing whether it was localized (specifically, to the hip or lumbar spine), or generalized. The study found a general loss of bone density through DXA scans and also a reduction in lower limb muscle strength [69]. A number of reported cases, have also found that apart from loss of bone density, osteonecrosis or osteomyelitis can be found in the jaw [70], as well as maxillary sinusitis [71]. Any effects on dental and oral health are reported in a separate section below.

3.1.4. Cannabinoids

Cannabinoids are the chemical components found in the Cannabis plant (Marijuana). The main psychoactive chemical is tetrahydrocannabinol (THC). The drugs can be smoked, inhaled or eaten. Cannabis (marijuana) or hemp are legally accepted in some regions and countries as they also demonstrated health benefits [72]. Indeed, the chemical components activate the endocannabinoid receptors of the body and brain resulting in a feeling of happiness, but they can also affect bone homeostasis [72,73] (http://www.thedrugswheel.com/; https://adf.org.au/drug-facts/cannabinoids/ (accessed 29 October 2021)). Studies have shown a significant decrease in bone mass density and bone quality among smokers of marijuana with respect to non-smokers [74]. Paradoxically, depending on the age of the individual, cannabis can also help with bone loss and has been used to manage osteoporosis [75]. However, no correlation was found between cannabis consumers and bone density in a study on the femur and lumbar spine in a U.S. study [76]. The positive and negative effects are still unclear at present [72,77]. The effects of Marijuana on teeth is covered in a separate section below.

3.1.5. Alcohol

Alcohol is a depressant like diazepam or benzodiazepines, thus slowing down the message between the brain and the body, and hence its vital functions. Depending on the amount taken and body composition, however, it can also act as a stimulant. A number of publications have examined the association between alcohol and bone disease in adolescents and adults [78–80]. The effects of light consumption, long-term and binge-drinking have been investigated in clinical studies [79]. It has been demonstrated that alcohol can affect bone proliferation and lead to low bone density (leading to osteopenia and possibly osteoporosis) and strength due to a remodeling imbalance [81–83]. However, this is dependent on the pattern of consumption and intake [84,85]. One study revealed that 12% of fractures in middle-aged men, could be avoided if alcohol, as well as smoking, were eliminated [86]. Alcohol can also inhibit osteoblast proliferation and thus be detrimental to fracture healing [87]. One paper in forensic anthropology suggested that an individual's age-at-death may have been overestimated from the skeletal remains of a person who suffered from alcoholism. The case presented cortical thinning, 'light' bones, as well as various skeletal fractures in different stages of healing; although these characteristics may more likely be secondary to alcoholism than due to the age of the individual [20]. Furthermore, osteonecrosis associated with alcoholism has been identified and widely reported in the clinical literature, especially avascular necrosis of the femoral head [88,89]. Much information is also available relating to alcohol and pregnancy, which is not covered in detail here, but it is worth mentioning a number of skeletal anomalies affecting cranial suture such as craniosynostosis in the fetus due to alcohol consumption during pregnancy [90].

3.1.6. Tobacco

There has been much research on the impact of smoking (nicotine and tobacco) on health, some of which has focused on bone health [91,92]. Amongst the skeletal complications caused by smoking are lower BMD [93,94] although this is still debatable [95–97], higher fracture risk [97], and delayed bone fracture healing and further complications [98–100]. A study on young adult (18-19 years) men, smokers vs. non-smokers, showed a reduction in BMD and also reduced cortical thickness in radius and tibia [101]. This in turn leads in smokers to an increase in fractures, especially osteoporotic fracture sites such as the spine, hip, wrist or major long bone shafts, but not to the skull [86]. Scolaro et al. [102] further demonstrates complications with fracture healing and nonunion in some instances. This delayed healing may be related to poor bone mineralization and smoking impairing Type I collagen fibrils [103] as well as other factors [104]. Complications of smoking on oral health are explored later, as well as in cartilage [105,106]. Pathological conditions may also be considered as a result or in association with tobacco, for instance an increase in degenerative joint disease in the vertebrae [107] or children in a smoking intrauterine and post-uterine environment where their skeletal growth and development may be affected [108].

3.1.7. Oral Glucocorticoids

Glucocorticoid-induced osteoporosis is the most common iatrogenic cause of secondary osteoporosis. The direct effect that this class of drugs has on the skeletal structure is drug-induced osteoporosis if used long-term [109]. These drugs also affect the endocrine system, which controls a number of different hormone mechanisms, causing disorders such as hypogonadism, which again can affect bone turnover and decrease BMD [110]. Glucocorticoids are a class of corticosteroids, which regulate the metabolism of glucose in the body and are widely used in the medical sector for conditions that are caused by inflammation, such as asthma, allergies, auto-immune diseases and sepsis [111]. Prolonged or incorrect use of these can result in osteoporosis, osteonecrosis, high fracture risk and slower fracture repair [109,112]. Slower fracture repair especially callus formation and healing has also been observed in mice [113]. In children, glucocorticoids will result in short stature, delayed growth and maturation, unless reversed with growth hormone therapy [114,115]. This delayed growth can occur within three months of treatment with glucocorticoids and skeletal deformity may result from long-term treatment in children [116]. It may delay carpal bone age as observed in a Chinese study [117], a consideration relevant if estimation age in the living [118].

3.1.8. Non-Steroid Anti-Inflammatory Drugs (NSAIDs)

Non-steroidal anti-inflammatory drugs (NSAIDs) are some of the most prescribed medications worldwide, with analgesic, anti-inflammatory, antipyretic and platelet antiaggregant functions [119,120]. This heterogeneous group of drugs acts by blocking cyclooxygenase enzymes (cox-1 and cox-2), which in turn inhibits prostaglandins synthesis, which has an important role in bone turnover by influencing both osteoblast and osteoclast activity [121–123]. Several studies have explored the effects of NSAIDs on fracture healing, as these drugs are commonly used for fracture and postoperative pain control following orthopaedic surgery [124,125]. However, some of these studies report how NSAIDs may delay bone healing [119,126–130]. An increased incidence of nonunion fractures, malunion and infections are observed, with examples of case reports of this in the femoral shaft and the spine [120,125,131–134]. However, some of this data has been extrapolated from animal studies, while human trials have not always reported strong evidence of this association [87,124,135,136]. NSAIDs also seem to impair entheses (tendon-to-bone) healing [123] and accelerate cartilage degeneration in osteoarthritis [137,138]. Regarding skeletal trauma, not all NSAIDs have been found associated with an increased risk of fractures [139]. For instance, diclofenac and naproxen have been associated with an increase fracture risk in hip, spine, and forearm; while others showed either a higher BMD, with a potentially lower fracture risk [136,140]; or did not show any association (e.g., aspirin) [119,141–143]. This

positive effect on BMD (total and hip) was observed with increasing doses, whereas it decreases at low doses, potentially increasing the fracture risk [134,139,144]. In paediatrics, no effects on bone have been reported on low dose and short duration therapy [129,145]. By contrast, if chronically prescribed during pregnancy, and depending on the gestation period, NSAIDs may have adverse skeletal effects on the fetus and newborn, including presence of cleft palate, decreased skeletal development, decreased vertebral and fracture callus mineralization, decreased fetal length, fused ribs, incomplete ossification of the cervical arch, deformation of lumbar arch, and absent sacral arch [146].

3.1.9. Paracetamol

Paracetamol (acetaminophen) is a drug with analgesic, antipyretic and mild anti-inflammatory properties, and is one of the most used medications worldwide [147]. Its mechanism of action involves the cyclooxygenase (COX) and cannabinoids pathways, decreasing prostaglandins production and in turn affecting bone turnover [140]. However, despite its very wide usage, very few studies have explored the potential link between this drug and bone health [148]. Changes in BMD and bone fragility with an increased risk of fractures have been the most studied [143]. Several authors have reported no difference in BMD between paracetamol users and non-users [147,149]. No significant differences were found according to dose and pattern of users (intermittent vs. continuous) [143]. By contrast, other studies have shown a decrease in BMD over time, although smaller than other analgesics such as NSAIDs and opioids [150]. Similar results are found when investigating the risk of fractures [139,140,143]. The risk of fracture has been reported for the spine, hip, and forearm, and it is not dose-dependent [139]. Moreover, the effects of this drug on proliferation and differentiation of osteoblasts, if used in the early phases of healing, may impair bone regeneration and implant osseointegration [148]. In contrast, other studies have not supported this association, for example Vestergaard et al. [143] detected slightly higher levels of alkaline phosphatase, a marker of bone turnover. Since conflicting results have been found so far on the effects of paracetamol on bone, and little is known about them [143], further studies are therefore needed to better investigate and understand the impact of this drug on bone health [140].

3.1.10. Gonadotropin Releasing Hormone Agonists (GnRHa)

Gonadotropin releasing hormone agonists (GnRHa) are commonly used for treatment of several conditions, including breast cancer, prostate cancer, endometriosis, gender dysphoria and central precocious puberty (CPP) [151,152]. They act on the pituitary-hypothalamic-gonadal axis inducing secondary hypogonadism and reducing the production of sex steroid hormones in both sexes, oestrogens in women and androgens in men [152,153]. These hormones influence osteoblasts and osteoclasts activity, with important functions in bone turnover including bone growth and maturation [154,155]. Due to sex hormones deprivation, bone turnover is accelerated with suppressed bone formation and increased bone resorption. Therefore, GnRHa may have a detrimental effect on bone health causing reduction of BMD and increasing the risk of osteoporosis and fractures, as reported by several studies [153]. GnRHa are extensively used as adjuvant endocrine therapy in breast and prostate cancer [152], leading to a cancer treatment-induced bone loss [154]. This accelerated bone loss involves trabecular bone (spine) and is greater in women than in men ([152], resulting in a BMD reduction estimated between 5% and 10% in spine and hip after one year, and continuing to decrease in long-term therapy ([153]. GnRHa therapy also increases the risk of osteoporosis and fractures, with a longer therapy duration and a higher number of doses predicting a greater risk [156,157]. In women, lumbar spine and femoral neck fractures are the most commonly affected. In men, the radius, vertebra and hip/femur are the most frequently fractured bones [152]. GnRHa have been used to reduce pelvic pain, but this in turn has shown to lead to a reduction in BMD in the lumbar spine, hip/proximal femur and radius after 6 months of treatment, sometimes followed by a partial or complete recovery after a withdrawal of 6 months-1 year [155,158–161]. Differences have also been

observed between different GnRHa, with leuprolide acetate having a greater detrimental effect on BMD than buserelin for example [155]. Whilst short-term therapy would unlikely cause bone loss, little data is available on the long-term consequences with regard to low BMD and fracture risk [155]. These drugs are also used in gender dysphoria and CPP in children and adolescents. The effects on bone are of concern due to the hormonal suppression occurring in puberty [162], potentially delaying or attenuating peak bone mass (PBM) although this is still not fully understood [163]. A decrease in BMD was observed in lumbar spine and femoral neck in transgender individuals [163,164] as well as in CPP, but with the latter showing reversible effects after withdrawal [151,165]. Nonetheless, attaining a normal PBM does not seem to be impaired [162].

3.2. Proton Pump Inhibitors

Proton pump inhibitors (PPI) are considered relatively safe and are widely used as acid-suppressor medicine to treat acid-related diseases (e.g., gastroesophageal reflux, peptic ulcers, heart-burn, dyspepsia, chronic cough, prevention of gastric injuries from NSAIDs and surgery) [166].). They are a class of drug that act on the cells that line the gastrointestinal tract and reduce acid production, allowing the lining to heal, or to prevent an ulcer from occurring [167]. There is a large body of evidence that demonstrates an association between PPI therapy and risk of fractures, in particular a moderate increased risk of any fractures in particular to the hip and spine, with a stronger association of hip fractures with increased duration of PPI treatment, as well as an association between PPI therapy and osteoporosis [166,168]. The association between PPI use and BMD is debatable, with some studies showing BMD loss [169] and others concluding an absence of correlation [166,170]. Two main factors may explain the association between PPI therapy and increased fracture risk as well as osteoporosis. Firstly, decreased calcium absorption has been noted in patients taking PPI, which would cause an increased rate of bone resorption; however, there are various factors, which may influence calcium absorption (e.g., dietary calcium intake and time of medication) [166]. Secondly, a selection bias and the absence of adjustment for cofounders (which include a large number of comorbidities and medication): older and sicker patients tend to be treated with PPI, and frailty and old age are risk factors for fractures [166,168].

3.3. Antiretroviral Therapy

Antiretroviral therapy (ART) are drugs that are taken to treat and prevent mortality and morbidity by retrovirus infections, such as human immunodeficiency virus (HIV). These drugs help control the virus by lowering the viral load, preventing transmission, and increasing life-expectancy rather than actually curing the disease [171]. Whilst there may be about a dozen drugs to treat HIV, it is a combination of these that are prescribed for therapy. HIV is already known to affect the skeletal system through low BMD, osteoporosis, osteonecrosis and more rarely, osteomalacia, as well as fractures and HIV-induced infections and inflammations [172–174]. Osteonecrosis is commonly present in the proximal femora and may be bilateral [175,176]. (Regarding ART, several studies have demonstrated an association between long-standing ART and lower BMD in HIV individuals [173,174,177–179], although other research reported no determining effect of ART on BMD [180]. Overall, low BMD in HIV patients results from a multifactorial interaction between HIV infection, conventional risk factors for osteoporosis, ART-related complications and HIV/AIDS-related conditions (e.g., muscle wasting, kidney disease, vitamin D deficiency and hypogonadism) [177,181–183]. In addition to low BMD, both long-standing HIV and ART have been reported to be associated to osteopenia, osteoporosis, osteonecrosis, osteomalacia and a higher rate of fractures [173,179]. Indeed, ART has a direct effect on the bone metabolism by exacerbating bone loss (with a reported 2–6% loss in BMD) at the femora, lumbar spine, and hips; which are sites susceptible to fractures [173,179,183]. Lastly, neuropathy may be another potential complication of ART [184,185], which may indirectly impact the skeletal system by leading to conditions such as neuropathic arthropathy (Charcot joint) [186,187].

3.4. Anti-Depressant Drugs

Patients that suffer with depression often have low levels of serotonin, which is a neurotransmitter found mainly in the gastrointestinal tract, platelets and the central nervous system (CNS) and is a contributor to feelings of wellbeing and happiness (InformedHealth.org (internet). Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2006. Depression: how effective are antidepressants? (Updated 18 June 2020) (accessed October 2021)). In some countries they are the most used therapeutic medications [188]. It also regulates the skeletal response to parathyroid hormone due to its receptors that are found on osteoblasts and osteocytes. Two commonly prescribed classes of drugs are selective serotonin re-uptake inhibitors (SSRIs) and tricyclic anti-depressants (TCAs) [189]. This paper focuses mainly on SSIRs, which seem to be associated to bone metabolism [190,191]. Bone loss density, rapid bone loss in certain age groups and an increase in osteoporosis in men has been shown in those taking anti-depressant drugs [192–194] and thus a risk of osteoporotic fracture [195,196]. Furthermore, in an experimental animal study, sertraline was shown to impair and disrupt bone healing with significant decrease in trabecular thickness ([197,198].

The link between fracture risk and SSIRs has been widely noted, however, depression itself has been shown to correlate with a decreased bone mineral density and increase fracture risk [199]. Although, taking into consideration the psychological condition of the individual receiving treatment, there is a likely chance that there will be other lifestyle risk factors, which may influence bone mineral density and increased fracture risk, such as smoking, increased alcohol consumption and physical inactivity [189]. Thus further work is required to show any link with depression, drugs and bone health [193,200].

3.5. Anti-Epileptic Drugs

Chung & Ahn [201] discuss the effects of anti-epileptic drugs (AED) and their effect on bone in children being treated for epilepsy. The authors examined bone density scans on a number of skeletal areas including the upper and lower limbs, the ribs, pelvis, and spine in a sample of 78 epileptic and 78 control patients, and concluded that the former group, which was treated with AED, had lower bone density. Lower bone density in those taking AED seems to correlate in other studies for different anatomical regions [202–204]. Other studies in adults have shown no known significant differences between short-term and long-term use of these drugs in the overall skeleton, but significant differences when specific bones are taken into account, such as the tibia and innominates [205]. It has been suggested that the reason for this lower BMD is that anti-epileptic drugs directly inhibit osteoblast function as well as inhibiting intestinal calcium absorption [109]. This reduction in bone mass density also increases fracture risk. In adults, the association with osteopenia and osteoporosis has been demonstrated [206,207], with increased fracture rates associated to the drugs as well as the result of seizures. Although the results are conflicting [208], generally speaking these drugs will lead to low bone density as well as an increased risk of fractures [209,210]. Reduced levels of Vitamin D have also been observed with AED intake [211,212] and also retarded growth and stunting [213].

3.6. Antidiabetic Drugs

These medications, including insulin, exist to control and maintain glucose or sugar levels in the blood and thus more commonly used to manage diabetes, adversely affect bone metabolism [214], especially by impairing osteoblast function and activating osteoclastogenesis [215]. This may ultimately lead to a decrease in low bone mineral density, decrease bone strength related to low bone turnover, alteration of the microstructure, and a risk of osteoporotic fractures such as at the hip [216,217]. This is of course also drug type dependent [218,219]. For example, thiazolidinedione in particular is associated with secondary osteoporosis and an increased fracture risk [219–221]. Overall, antidiabetic drugs are linked to an increase risk of osteoporosis, fractures and possibly osteoarthritis too [218,222,223], although this latter is not yet clear [224].

3.7. Antiresorptive Drugs

These drugs include a class termed bisphosphonates. These inhibit osteoclastic activity and although bisphosphonates are likely to control osteolysis in tumors and disease progression [225,226], they also do have other effects, for instance osteonecrosis of the jaws and more frequently in the mandible [227,228]. Osteonecrosis of the jaw (ONJ) is a well-known complication of antiresorptive or antiangiogenic therapy for the management of osteoporosis and other cancer-related conditions [229]. Available data indicate that 5% of patients exposed to antiresorptive agents may develop ONJ, depending on the duration of therapy. Oral surgical procedures, tooth extractions and infection of the mandible and/or maxilla are considered the main risk factors for developing ONJ when receiving antiresorptive therapy [228]. A study by Gupta and Gupta [227] indicates that osteonecrosis tends to develop in the jaw because it has a higher remodeling rate than other bones, making it more prone to the effect of bisphosphonates. The three most common sites for ONJ are (1) nonhealing dentoalveolar sites or dental extraction sites; (2) traumatized tori (palatal and/or mandibular); and (3) exposure of portions of the mylohyoid bridge [227,230,231]. Osteomyelitis and abscesses may also be present and in living individuals exposed bone too [231,232].

Bisphosphonates with denosumab are the most commonly used antiresorptive drugs and although they cause osteonecrosis of the jaw [233] when used to treat malignant disease, they are used to treat osteoporosis and the risk of fracture associated from it [234,235].

3.8. Antithrombotic Drugs

Antithrombotic drugs can be antiplatelets (e.g., aspirin) or anticoagulants (e.g., heparin, warfarin) and prevent blood clots from forming. A number of groups would appear through a literature review to affect bone health, primarily linked to osteopenia [236]. Some anticoagulants such as heparin may result in lower bone mass density, influencing bone metabolism and resulting in an increased risk of osteoporotic fractures [237,238]. Impaired fracture healing may also take place [239]. One study on warfarin demonstrated an association with a decrease in BMD in the calcaneus of patients compared to non-patients through examination with a quantitative ultrasound [240]

3.9. Other Drugs

This paper has not covered all drug groups, all the different classes of drugs, nor the combination of taking several classes together and how this may affect the skeleton. Nevertheless, it is worth mentioning here a number of other drugs that may have a significant effect on the skeleton too. For example, Depot Medroxyprogesterone Acetate (DMPA) or Depo-Provera is a contraceptive drug that both adult and adolescent females may take. It works by inhibiting luteinizing hormone (produced and released by the pituitary gland) and follicle stimulating hormone (also released from the pituitary gland and important for the reproductive system in men and women), which in turn decreases oestrogen production [118], thus resulting in decreased bone mineral density and increase risk of osteoporosis during its use. Most bone loss occurs in the first two years of use and mainly seen in the vertebral column and hips – and so this is where most fractures are seen. A study conducted using a group of physically active female army recruits indicated that there was a marked increase in stress fractures in the calcaneus of the individuals taking DMPA [109].

Amongst the hormone therapy drugs, aromatase inhibitors (AI) which has been used to treat a number of diseases such as breast cancer, does result in bone loss and a risk of fractures [241,242]. Another hormone treatment is thyroid hormone therapy (THT), which is used to compensate for an underactive thyroid. Patients with hypothyroidism undertaking THT may or may not see skeletal changes. The literature provides conflicting reports on BMD, with some studies showing BMD loss while others found no changes [243–245].

Antineoplastic drugs are chemotherapy drugs and highly toxic but used to treat cancer. Since there are almost 2000 medications under this class of drug (https://seer.cancer.gov/

seertools/seerrx/, accessed 31 October 2021), it is impossible to cover here, especially when in different forms. It is worth indicating that some side effects will include lower BMD, bone marrow suppression, haematological complications including anaemia, periapical lesions possibly leading to osteomyelitis, etc. [246–250].

It is also worth mentioning antipsychotic medications or agents, used to manage and/or treat patients with psychosis (e.g., in patients with schizophrenia, bipolar disorders, etc). Antipsychotic drugs have a physiological effect on bone, as they increase the concentration of prolactin, which lowers oestrogen and testosterone levels potentially leading to bone loss [251]. One study indicated that the risk of a hip fracture was increased 5-fold in older women and 6-fold in older men taking antipsychotic drugs [109]. In one study, it was also noted that in young men and pre-menopausal women these drugs lower bone mineral density as much as 20% in the spine [252].

3.10. Oral Pathology

Although already introduced above, it is worth providing an overview of the dental and oral (bone) complications that can arise in patients taking some of these drugs. In forensic cases, this is the remit of the forensic odontologist, but nevertheless it is important for forensic anthropologists to be aware of these changes, in addition to other factors that may affect oral pathology such as lifestyle or oral hygiene practices.

Tomita et al. [67] indicate that in a very short space of time, rampant caries is often found in methamphetamine users ("meth mouth"). In addition to caries, periodontal disease and tooth loss [253], partly due to the reduction of saliva in the mouth and other factors [254]. Cocaine can also damage teeth and the surrounding soft tissue, as one common method for consumption is by rubbing the powder against the gums or gingivae. Cocaine reduces salivary pH leading to dental erosion, and there is a higher risk of periodontal disease and tooth loss [255–259]. Smoking or eating cannabis has also similar effects [260].

Tooth discoloration can be caused by medication such as antihistamines and antibiotics amongst others [261–263]. Furthermore, enamel hypoplasia as well as microdontia and hypodontia can be found in children treated with antineoplastic drugs [250]. Smoking tobacco can also cause tooth discoloration [264] although there are many other factors influencing color staining in teeth [265].

With regard to the alveolar bone and further involvement of some of the drugs included above such as anti-resorptive drugs and antineoplastic drugs can result in osteonecrosis of the jaw [63,64].

3.11. Other Skeletal Involvement

As the skeleton is also composed of cartilage and cartilage degeneration will affect some of the indicators forensic anthropologists use in reconstructing the biological profile, it is necessary to point out how some drugs and medications can affect cartilage. For example, smoking tobacco has been found to be associated with cartilage loss and defects in the cartilage of offspring [106]. This association resulting in osteoarthritis has also been proven in other studies. Amin and colleagues [105] found that men with knee osteoarthritis who smoke sustain greater cartilage loss and have more severe knee pain than men who do not smoke. One area of interest may be the calcification of cartilage, whether costal (sternal end of the rib) or any other (e.g., thyroid cartilage). Premature calcification of cartilage has been attributed to a number of aetiologies [266]. However, chondrocalcinosis as well as chondritis has also been attributed to certain drugs such as corticosteroids, bisphosphonates and others [267–270].

As forensic anthropologists during recovery of human remains, we may encounter urinary or renal stones or calculi. These can also be drug induced [271,272].

3.12. Summary of Results and Further Observations

Taking all the above information gathered from an exhaustive literature search in the medical, clinical, pharmacological and other disciplines; it can be understood that

medication and the abuse of drugs can have an effect on the skeleton. These include loss of bone density often leading to osteoporosis and risk of fractures, necrosis, joint disease, delayed maturation, delayed fracture healing, cartilage calcification, and oral pathologies (Table 1). Whist some of these drugs may affect the skeleton generally, some studies have focused on particular regions or elements and some medications definitely only involve certain areas (Table 1), such as the vertebrae, long bones, mandible or maxilla. These may influence the estimation of the biological profile related to age-at-death estimation, sex estimation and other parameters or features used to identify the deceased. Table 1 should assist forensic anthropologists in their awareness of how certain medical histories and the associated use of certain drugs may affect the skeleton. This may be an important consideration when reconstructing the biological profile. In addition, some of these skeletal alterations may reflect a person that was using certain medication(s) and thus it may also be able to help with narrowing down the list of missing people.

Table 1. Summary of the effects of drugs on bone, as taken from the literature review for this paper. For references or bibliography, see the main body of text.

Drug		Effect on Bone	Location
Cocaine		Cocaine-induced midline destructive lesion (CIMDL) and other nasal deformities, septum perforation, infection (e.g., maxillary sinusitis) Periodontitis, dental caries, (ante-mortem) tooth loss, dental erosion.	Nasal septum, nasal walls, hard palate, maxilla and orbital walls. Dentition.
Opioids	Morphine	Osteoporosis, osteopenia, increase risk of fracture, longitudinal growth, skeletal development.	Not specific. Some fractures may be at sites such as hip, spine, forearm but not always attributed to osteoporosis. Cartilage affected during growth and development.
	Methadone	Increased risk of osteoporosis and osteopenia, increased risk of fracture, decrease in bone mineral density.	Not specific. Some fractures may be at sites such as hip, spine, forearm but not always attributed to osteoporosis.
	Heroin	Decrease bone mineral density, osteoporosis, osteopenia, septic arthritis, bone turnover, osteomyelitis.	Not specific. Septic arthritis in sacroiliac, costoclavicular, hip and shoulder joint. Sometimes osteomyelitis in long bones at sites where injections.
Amphetamines		Osteonecrosis, Osteoporosis, Osteopenia, loss of bone density, maxillary sinusitis, osteomyelitis 'Meth mouth': Dental caries, periodontal disease, tooth loss, periodontitis, dental caries, dental erosion.	Loss of bone density throughout body. Osteonecrosis of jaw. Sinuses Dentition ('Meth mouth').
Cannabinoids		Possible loss of bone density, leading to osteoporosis and increased fracture risk. Periodontal disease.	Not specific. Dentition.
Alcohol		Effect on osteoblast proliferation, lower bone density, osteopenia, osteoporosis, increased fracture, poor fracture healing, avascular necrosis.	Throughout skeleton. This effect may depend on sex, age and lifestyle factors, patterns of drinking, volume of alcohol, etc. Avascular necrosis of femoral head.
Tobacco		Bone density, bone fractures, delayed haling of fractures or non-union. Periodontitis.	Throughout skeleton. Sites of osteoporotic fractures. No fractures to skull. Dentition.
Oral Glucocorticoids		Increased risk of osteoporosis, decrease in bone mineral density, fracture risk, slow fracture healing, delayed maturation, short stature.	Not specific.

Table 1. *Cont.*

Drug	Effect on Bone	Location
Non-steroidal anti-inflammatory drugs (NSAIDs)	Delayed fracture and entheses healing. Fracture nonunion/malunion. Possible cartilage degeneration. Increase/decrease in BMD (type and dose-dependent). Possible increased/decreased fracture risk. Possible skeletal effects in fetus and newborn (therapy during pregnancy).	Changes not specific, observed hip, femur, spine, and forearm In fetus/newborn—cleft palate, fused ribs, decreased vertebral mineralization, deformation of lumbar arch, absent sacral arch, incomplete ossification of cervical arch, absent/hemicentric body of thoracic or lumbar vertebra.
Paracetamol	Possible decrease in BMD. Possible increased risk of fractures (at low doses). Possible impairment of implant osseointegration.	Observed in spine, hip, forearm.
Gonadotropin-releasing hormone (GnRH) agonist	Decrease in BMD (potentially reversible after treatment). Increased risk of fractures. Osteoporosis. Possible delay/attenuation of PBM.	Trabecular bone (lumbar spine), but also observed in hip, proximal femur, and radius.
Proton pump inhibitors	Increased risk of fractures. Osteoporosis. Possible decrease in BMD.	Any site, but in particular at the hips and lumbar vertebrae.
Antiretroviral therapy	Decrease in BMD, osteopenia, osteoporosis, osteonecrosis, osteomalacia, increased risk of fractures. Charcot joint (indirectly).	Throughout the skeleton, particularly at the femora, lumbar vertebrae and hips. Osteonecrosis on proximal femora, sometimes bilateral.
Antidepressant drugs	Reduced estrogen production. Osteoporosis. Increased risk of fracture. Decrease in bone mineral density.	Throughout skeleton. Osteoporotic fracture sites.
Anti-epileptic drugs	Decrease in bone mineral density and osteoporosis, increased risk of fracture, retarded growth and stunting.	Throughout skeleton.
Antidiabetic drugs	Decrease in bone mineral density, alteration of bone microstructure, increase risk of fractures, possibly osteoarthritis.	Throughout skeleton but increase risk in fracture particularly related to osteoporotic fracture sites.
Antiresorptive drugs	Osteonecrosis of the jaw.	In particular the mandible.
Antithrombotic drugs	Decrease in bone mineral density, increase risk of fractures and impaired fracture healing.	Throughout skeleton. Fractures at osteoporotic fracture sites.

Table 2 summarizes those drugs that particularly lead to loss of bone mineral density, potentially osteoporosis and risk of fracture. An additional column for bone destruction as also been included.

In addition, it is worth stating that apart from knowing the effects of drugs on bone there is potential to investigate these post-mortem through toxicological analysis of the bone. As drugs can be incorporated into bone through superficial arteries, born in the periosteal network, which later diffuse into the peripheral layer of the compact bone; or they can circulate through deep arteries and nutrient foramina toward the spongy bone tissue to terminate in the bone trabeculae and bone marrow; within the bone matrix, drugs can remain for instance in hydroxyapatite and be incorporated into the inorganic matrix through bone remodeling. Through these mechanisms, drugs can be preserved and detected in bone tissue even after a long post-mortem interval [18,27]. As evidenced in Table 3, toxicological analyses have been performed on various bone samples such as the cranium, rib, femur, vertebra, clavicle, and iliac crest [27,273,274]. As a result, the literature (Table 3) reports the detection of MDA (amphetamine), ketamine (anesthetic), pregabalin and carbamazepine

(anticonvulsant drugs), diphenhydramine (antihistamine drug), atenolol and bisoprolol (antihypertensive drugs), caffeine, cocaine and its metabolite (stimulants), THCCOOH (metabolite of THC, a cannabinoid), laudanosine (metabolite of atracurium, a curare), as well as many antidepressants, antipsychotic drugs, benzodiazepines and opioids.

Table 2. Summary of effects on bones according to the drug class discussed in this paper. The absence of any information in the cells does not necessarily mean that these changes do not occur in a particular drug, but it has not been noted in the literature consulted for this paper. Code: Y = yes.

	Decreased BMD/Osteoporosis	**Increased Risk of Fractures**	**Bone Destruction/Osteonecrosis**
Cocaine			Y
Methadone	Y	Y	Y
Heroin	Y	Y	Y
Amphetamines	Y		Y
Cannabinoids	Y?		
Alcohol	Y	Y	Y
Tobacco	Y	Y	
Oral glucocorticoids	Y	Y	
NSAIDs	Possibly Y (low doses)	Possibly Y (when BMD is decreased)	
Paracetamol	Possibly Y	Possibly Y	
GnRH agonist	Y	Y	
Proton pump inhibitors	Y	Y	
Antiretroviral therapy	Y	Y	Y
Antidepressant drugs	Y	Y	
Anti-epileptic drugs	Y	Y	
Antidiabetic drugs	Y	Y	
Antiresorptive drugs			Y
Antithrombotic drugs	Y	Y	

Table 3. Table that summarizes substances so far detected in bone through toxicological analyses in different studies. The table lists the substances, the study, the site of sampling, and the number of skeletons analyzed in the study.

Class of Molecules	**Drugs**	**Bone Samples**	**Number of Individuals Analyzed**	**Reference**
Amphetamines	MDA	Cranium	7	[27]
Anesthetics	Ketamine	Cranium, rib	19	[190]

Table 3. *Cont.*

Class of Molecules	Drugs	Bone Samples	Number of Individuals Analyzed	Reference
Anticonvulsant drugs	Pregabalin	Rib	3	[189]
	Carbamazepine	Femur	36	[275]
Antidepressants	Amitriptyline	Iliac crest, vertebra	39	[22]
		Femur	36	[275]
		Femur	6	[23]
		Rib	7	[276]
	Citalopram	Cranium, rib	19	[190]
		Iliac crest, vertebra	39	[22]
	Dothiepin	Femur	36	[275]
	Doxepin	Femur	36	[275]
	Duloxetine	Rib	7	[276]
	Mianserin	Femur	36	[275]
	Moclobemide	Femur	36	[275]
	Nordoxepin (Metabolite of doxepin)	Femur	36	[275]
	Nortriptyline	Femur	36	[275]
	Trazodone	Cranium, rib	19	[190]
	Venlafaxine	Cranium, rib	19	[190]
		Rib	7	[276]
Antihistamine drugs	Diphenhydramine	Iliac crest, vertebra	39	[22]
Antihypertensive drugs	Atenolol	Rib	2	[277]
	Bisoprolol	Rib	2	[277]
Antipsychotics	Chlorpromazine	Femur	36	[275]
	Clozapine	Femur	36	[275]
	Haloperidol	Cranium, rib	19	[190]
	Mesoridazine	Femur	36	[275]
	Promazine	Cranium, rib	19	[190]
	Quetiapine	Cranium	19	[190]
		Rib	3	[189]
	Thioridazine	Femur	36	[275]
Benzodiazepines	Alprazolam	Cranium, rib	19	[190]
	Bromazepam	Femur	6	[23]
	Delorazepam	Vertebra, rib	7	[27]
		Cranium, rib	19	[190]
	Diazepam	Cranium vertebra, rib	7	[27]
		Cranium, rib	19	[190]
		Iliac crest, vertebra	39	[22]
		Femur	36	[275]
		Femur	6	[23]
	Flurazepam	Cranium, rib	19	[190]

Table 3. *Cont.*

Class of Molecules	Drugs	Bone Samples	Number of Individuals Analyzed	Reference
	Lorazepam	Cranium	7	[27]
		Cranium, rib	19	[190]
	Lormetazepam	Cranium, rib	19	[190]
	Nordiazepam	Vertebra	7	[27]
		Cranium, rib	19	[190]
		Iliac crest, vertebra	39	[22]
		Femur	36	[275]
		Femur	6	[23]
	Oxazepam	Femur	36	[275]
	Temazepam	Femur	36	[275]
Cannabinoids	THCCOOH (Metabolite of THC)	Rib	7	[27]
Curare	Laudanosine (Metabolite of atracurium)	Iliac crest, vertebra	39	[22]
Opioids	6-MAM	Rib	6	[278]
	Buprenorphine	Vertebra	7	[27]
	Codeine	Iliac crest, vertebra	39	[22]
		Femur	36	[275]
		Femur	6	[23]
		Clavicle	3	[279]
	Meperidine	Iliac crest, vertebra	39	[22]
	Methadone	Cranium vertebra, rib	7	[27]
		Rib	6	[278]
		Femur	36	[275]
	Morphine	Rib	6	[278]
		Femur	6	[23]
		Femur	1	[280]
		Clavicle	3	[279]
	Norpropoxyphene (Metabolite of propoxyphene)	Iliac crest, vertebra	39	[22]
		Femur	36	[275]
	Oxycodone	Iliac crest, vertebra	39	[22]
		Femur	36	[275]
	Propoxyphene	Iliac crest, vertebra	39	[22]
		Femur	36	[275]
	Tramadol	Cranium, rib	19	[190]
		Rib	6	[278]
Stimulants	Caffeine	Femur	36	[275]
	Cocaine	Cranium, rib	19	[190]
		Femur	6	[23]
	Benzoylecgonine (Metabolite of cocaine)	Vertebra, rib	7	[27]
		Cranium, rib	19	[190]
		Rib	6	[278]
		Iliac crest, vertebra	39	[22]
		Femur	6	[23]

Although further research is required, results have shown that these drugs can be detected years after death.

4. Discussion

The aim of this paper is to increase awareness for forensic anthropologists on the effects that medication and drugs can have on the skeleton. This awareness will help with any considerations in forensic practice but it also opens new avenues for research. Prior to discussing the specific implications for biological profile and personal identification, a number of limitations need to be highlighted first.

4.1. Limitations

One of the limitations to highlight is that many if not most of these medications or drugs have similar effects on bones, and rarely are any of these changes pathognomonic to a specific drug, let alone other factors that can influence the alteration to the skeleton. For example, many of the drug classes described above will result in lower bone mineral density, and increase risk of osteoporosis and osteoporotic fractures. In turn, some drugs induce osteoporosis, for example, but osteoporosis can also occur as a natural disease.

A second and important limitation is that this study has taken each class of drug separately. Whilst a person may take one specific medication, this paper has not considered a combination of different drug classes and its effect on the skeleton. For example, the consumption of opioids in addition to prolonged alcohol ingestion. Furthermore, the effects after drug intake must be examined in detail to assess how long before any effects are reversed. This is beyond the scope of this paper.

Another difficulty in interpreting bone changes possibly related to drugs is that these may be influenced by a number of variables, including dosage, method of administration, and duration of treatment. All of these will have a different effect on the skeleton. One such example is cocaine, which if snorted, can cause destruction of the palatine and nasal bones [30]. The biological age of the individual as well as sex may also influence the effects on the skeleton.

Similar to palaeopathology or pathological alterations to the skeleton, diagnosis will be reliant on bone preservation, bone condition, skeletal completeness, distribution of the lesions over the skeleton, if unilateral or bilateral, etc. In addition, if bone mineral density is to be observed it is likely that specific imaging techniques are necessary, rather than a direct visual assessments of the bones.

As with many of the other drugs, a full understanding of each drug and its relation to the skeleton is not always clear, and is often dependent on age, sex and lifestyle factors. Moreover, to be more relevant to the forensic anthropologist, a more specific description of the skeletal lesions may be required, for example detailed information on osteonecrosis of the jaw including shape and dimensions of lesion, etc.

4.2. Implications for Forensic Anthropology: Effects on Age-at-Death, Sex Estimation and Other Parameters

As there has not been a published study in forensic anthropology regarding specifically how these drugs affect age and sex estimation methods, no definitive answer can be given. However, having observed some of the effects on the skeleton with some drugs, it can be hypothesized that some of these are likely to affect the indicators for age-at-death and sex estimation. Importantly, there may be issues around age estimation in the living, especially around skeletal growth. For age-at-death estimation, costal cartilage or pubic symphysis morphology may have been affected. For instance, if the individual presents with osteoporotic bones but is otherwise young, drugs such as corticosteroids, glucocorticoids, aromatase inhibitors and Depo-Provera could have resulted in this decreased bone density. Sex hormones may alter some sexual diagnostic features with more gracile bones and smaller muscle attachments, thus analysis of sex could be estimated incorrectly.

4.2.1. Implications for Personal Identification

Alterations of bone mineral density, such as osteopenia and osteoporosis, and their consequent increased risk of fractures are the most common effects of drugs on bones

as reported by the literature. However, due to their non-specificity, it is not possible to directly link these bone changes to the use of particular therapy drugs or drugs of abuse, as they could also be the outcome of normal ageing, other pathological conditions (such as endocrine disorders, eating disorders, immobilization, marrow-related disorders, disorders of the gastrointestinal or biliary tract, renal disease, and cancer) [281] and traumatic events; all potentially unrelated to the consumption of drugs. Table 4 proposes some possible influences of drugs in the reconstruction of the biological profile.

Table 4. Awareness of how drugs could affect biological profile reconstruction in forensic anthropology.

	Possible Effects	Observations
Age-at-death	Delayed maturation, pre-mature (costal) cartilage calcification, pubic symphysis morphology, joint disease, osteoporosis, tooth loss.	Likely age overestimation in adults. May require imaging such as body CT scans. Moreover, similar effects when estimation the age of a living person. If anomalies in age indicators perhaps enquire re medication and lifestyle environment.
Sex estimation	Possible morphological changes in pelvis and skull.	Misdiagnosis. Research in transgender individuals required too.
Physical attributes (stature, ancestry or population affinity)	Morphological assessment of nasal area may be altered by drug abuse. Stunted growth.	Ancestry estimation, stature.
Unique features	Osteonecrosis of the jaw, dental problems, fracture patterns.	May be able to indicate some possible medications or be consistent with medication intake.

In addition, results from forensic toxicology as seen in Table 3, in conjunction with skeletal changes that may be drug related, could help identification by adding to the biological profile.

5. Conclusions

Given the number of people taking drugs (including drugs of abuse and prescription medication) today, the aim of this paper was to present the main drugs that according to the medical and anthropological literature consulted for this paper have the potential to affect the skeleton directly. Through an extensive literature review, the information was evaluated and extrapolated from a forensic anthropology perspective, considering the impact for the reconstruction of a biological profile when studying unidentified human remains; or at least increase an awareness of possible alterations of the skeleton due to drugs and medication. The list of drug categories included here is more generic and does not address particular names of drugs or brands, or venture in any detailed characteristics of any alteration. Nevertheless, it provides an awareness on how drugs can possibly influence age-at-death, sex, stature and ancestry or biological affinity estimation, amongst other traits such as pathological conditions.

A number of questions arise from this review. One is that further research could target how medication may be affecting particular landmarks or refining those bone characteristics (e.g., location, dimensions, unilateral vs. bilateral, etc.), which may be used for biological reconstruction in forensic casework. Second, is that it may be worth seeing in terms of research what medical history a deceased had and explore whether this may have left any traits observed on their remains. Furthermore, it may be possible to consider medical history during our analysis of the individual, with regard to medication, in particular if forensic anthropologists have a medical background, or in conjunction with forensic pathologists and odontologists. As stable isotopes are also being used in forensic casework, a more in depth understanding of bone turnover may be worth exploring.

Although these modifications on the skeleton are not uniquely specific to a given substance, they can suggest drug intake in the differential diagnosis. This is therefore an

important piece information to be considered in forensic anthropological practice as it may implement the biological profile with unique information and improve accuracy in the application of standard methods and the interpretation of results. However, more research is needed to characterize with precision the effects of drugs on bones, and to clarify their influence on anthropological methods, for instance through the examination of skeletons with a known medical history of drug use. The knowledge reviewed in this study may be used in support or as basis for further research in forensic anthropology, but also potentially in the medical and pharmacological fields for/such as data to be shared more specifically.

Author Contributions: Conceptualization, N.M.-G.; methodology, N.M.-G., and V.J.; formal analysis, N.M.-G., V.J., E.B., L.B.-G., G.G., D.D.C., and C.C.; investigation, N.M.-G., V.J., E.B., L.B.-G., L.A., N.V.P., G.G., D.D.C., and C.C.; writing—original draft preparation, N.M.-G., V.J., E.B., and L.B.-G.; writing—review and editing, N.M.-G., V.J., E.B., L.B.-G., L.A., N.V.P., G.G., D.D.C., and C.C.; supervision—N.M.-G. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cattaneo, C. Forensic anthropology: Developments of a classical discipline in the new millennium. *Forensic Sci. Int.* **2007**, *165*, 185–193. [CrossRef] [PubMed]
2. Komar, D.A.; Buikstra, J.E. *Forensic Anthropology: Contemporary Theory and Practice*; Oxford University Press: New York, NY, USA, 2007.
3. de Boer, H.; Obertová, Z.; Cunha, E.; Adalian, P.; Baccino, E.; Fracasso, T.; Kranioti, E.; Lefévre, P.; Lynnerup, N.; Petaros, A.; et al. Strengthening the role of forensic anthropology in personal identification: Position statement by the Board of the Forensic Anthropology Society of Europe (FASE). *Forensic Sci. Int.* **2020**, *315*, 110456. [CrossRef] [PubMed]
4. Márquez-Grant, N.; Roberts, J. Redefining forensic anthropology in the 21st century and its role in mass fatality investigations. *Eur. J. Anat.* **2021**, *25*, 19–34.
5. Christensen, A.; Passalacqua, N.; Bartelink, E. (Eds.) *Forensic Anthropology: Current Methods and Practice*, 2nd ed.; Academic Press: London, UK, 2019.
6. Marquez-Grant, N. An overview of age estimation in forensic anthropology: Perspectives and practical considerations. *Ann. Hum. Biol.* **2015**, *42*, 308–322. [CrossRef] [PubMed]
7. Christensen, A.M.; Passalacqua, N.V.; Bartelink, E.J. Age Estimation. In *Forensic Anthropology: Current Methods and Practice, 2nd ed*; Christensen, A.M., Passalacqua, N., Bartelink, E., Eds.; Academic Press: London, UK, 2019; pp. 307–349.
8. Garvin, H.M. Adult Sex Determination: Methods and application. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; Wiley-Blackwell: Chichester, UK, 2012; pp. 239–247.
9. Ousley, S.D. Estimating Stature. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; Wiley-Blackwell: Chichester, UK, 2012; pp. 330–334.
10. Hefner, J.T.; Ousley, S.D.; Dirkmaat, D.C. *Morphoscopic Traits and the Assessment of Ancestry*; Wiley-Blackwell: Chichester, UK, 2012; pp. 287–310.
11. Ousley, S.D.; Jantz, R.L. Fordisc 3 and Statistical Methods for Estimating Sex And Ancestry. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; Wiley-Blackwell: Chichester, UK, 2012; pp. 311–329.
12. Ross, A.; Williams, S. Ancestry Studies in Forensic Anthropology: Back on the Frontier of Racism. *Biology* **2021**, *10*, 602. [CrossRef]
13. Latham, K.; Bartelink, E.; Finnegan, M. (Eds.) *New Perspectives in Forensic Human Skeletal Identification*; Academic Press: London, UK, 2018.
14. Goodman, S.B.; Jiranek, W.; Petrow, E.; Yasko, A.W. The Effects of Medications on Bone. *J. Am. Acad. Orthop. Surg.* **2007**, *15*, 450–460. [CrossRef] [PubMed]
15. Imbert, L.; Boskey, A. Effects of Drugs on Bone Quality. *Clin. Rev. Bone Miner. Metab.* **2016**, *14*, 167–196. [CrossRef]
16. UNODC. World Drug Report. 2021. Available online: https://www.unodc.org/unodc/en/data-and-analysis/wdr2021.html (accessed on 1 October 2021).
17. United Nations. World Drug Report 2018: Global Overview of Drug Demand and Supply. Latest Trends, Cross-Cutting Issues. 2018. Available online: https://www.unodc.org/wdr2018 (accessed on 15 March 2020).
18. Rubin, K. The manifestation of cocaine-induced midline destructive lesion in bone tissue and its identification in human skeletal remains. *Forensic Sci. Int.* **2013**, *231*, 408e1. [CrossRef] [PubMed]
19. Passalacqua, N. Drug Use, Homeostasis, and The Estimation of Age at Death from Skeletal Remains. 2014. Available online: https://www.academia.edu/6036089/Drug_use_homeostasis_and_the_estimation_of_age_at_death_from_skeletal_remains (accessed on 15 March 2020).

20. Michael, A.R.; Bengtson, J.D. Chronic alcoholism and bone remodeling processes: Caveats and considerations for the forensic anthropologist. *J. Forensic Leg. Med.* **2016**, *38*, 87–92. [CrossRef] [PubMed]
21. Andronowski, J.M.; Cole, M.E.; Davis, R.A.; Schuller, A.; Tubo, G.R.; LaMarca, A.R.; Taylor, J.T. The Longitudinal Effects Of Prolonged Opioid Use On Cortical Bone Remodeling In A Rabbit Model: Part I—Intraskeletal Variability and Regional Differences Detected Via Micro-Computed Tomography. In Proceedings of the American Academy of Forensic Sciences 73rd Annual Scientific Meeting 2021, Seattle, WA, USA, 9 December 2021.
22. McGrath, K.K.; Jenkins, A.J. Detection of Drugs of Forensic Importance in Postmortem Bone. *Am. J. Forensic Med. Pathol.* **2009**, *30*, 40–44. [CrossRef] [PubMed]
23. Orfanidis, A.; Gika, H.; Mastrogianni, O.; Krokos, A.; Theodoridis, G.; Zaggelidou, E.; Raikos, N. Determination of drugs of abuse and pharmaceuticals in skeletal tissue by UHPLC–MS/MS. *Forensic Sci. Int.* **2018**, *290*, 137–145. [CrossRef] [PubMed]
24. Rubin, K.M. The current state and future directions of skeletal toxicology: Forensic and humanitarian implications of a proposed model for the in vivo incorporation of drugs into the human skeleton. *Forensic Sci. Int.* **2018**, *289*, 419–428. [CrossRef]
25. Smith, S.Y.; Doyle, N.; Felx, M. Introduction and Considerations in Bone Toxicology. In *Bone Toxicology*; Smith, S., Varela, A., Samadfam, R., Eds.; Springer: Cham, Germany, 2017; pp. 3–26.
26. Watterson, J. Challenges in forensic toxicology of skeletonised human remains. *Analyst* **2006**, *131*, 961–965. [CrossRef] [PubMed]
27. Giordano, G.; Biehler-Gomez, L.; Seneci, P.; Cattaneo, C.; Di Candia, D. Detecting drugs in dry bone: A pilot study of skeletal remains with a post-mortem interval over 23 years. *Int. J. Leg. Med.* **2021**, *135*, 457–463. [CrossRef]
28. Richards, V. An Overview of How Drugs Affect the Skeleton: Implications for Forensic Anthropology. Unpublished Master's Thesis, Cranfield University, Cranfield, UK, 2014.
29. Drug Enforcement Administration. Drugs of Abuse. 2017. Available online: https://www.dea.gov/sites/default/files/drug_of_abuse.pdf (accessed on 16 March 2020).
30. Siegrist, M.; Wiegand, T.J. Cocaine. In *Encyclopedia of Toxicology: Third Edition*; Wexler, P., Ed.; Academic Press: London, UK, 2014; pp. 999–1002.
31. Néel, A.; Agard, C.; Hamidou, M. Vasculitides induced by cocaine and/or levamisole. *Jt. Bone Spine* **2018**, *85*, 9–14. [CrossRef] [PubMed]
32. Tran, H.; Tan, D.; Marnejon, T.P. Cutaneous Vasculopathy Associated with Levamisole-Adulterated Cocaine. *Clin. Med. Res.* **2012**, *11*, 26–30. [CrossRef]
33. Deutsch, H.L.; Millard, D.R. A New Cocaine Abuse Complex: Involvement of Nose, Septum, Palate, and Pharynx. *Arch. Otolaryngol. Head Neck Surg.* **1989**, *115*, 235–237. [CrossRef]
34. Gupta, A.; Hawrych, A.; Wilson, W.R. Cocaine-Induced Sinonasal Destruction. *Otolaryngol. Neck Surg.* **2001**, *124*, 480. [CrossRef]
35. Padilla-Rosas, M.; Jimenez-Santos, C.I.; García-González, C.L. Palatine perforation induced by cocaine. *Medicina Oral Patología Oral y Cirugia Bucal* **2006**, *11*, E239–E242.
36. Schweitzer, V.G. Osteolytic sinusitis and pneumomediastinum: Deceptive otolaryngologic complications of cocaine abuse. *Laryngoscope* **1986**, *96*, 206–210. [CrossRef]
37. Molina, P.C.; Carmona, E.F.; Palza, C.A.M.; Serrano, R.L.T. Orbital and Nasal Complications Secondary to Inhaled Cocaine Abuse. *Acta Otorrinolaringol.* **2012**, *63*, 233–236. [CrossRef] [PubMed]
38. Lascaratos, G.; McHugh, J.; McCarthy, K.; Bunting, H. Advanced cocaine-related necrotising sinusitis presenting with restrictive ophthalmolplegia. *Orbit* **2016**, *35*, 164–166. [CrossRef]
39. Brembilla, C.; Lanterna, L.A.; Risso, A.; Bombana, E.; Gritti, P.; Trezzi, R.; Bonaldi, G.; Biroli, F. Craniovertebral junction instability as an extension of cocaine-induced midline destructive lesions: Case report. *J. Neurosurg. Spine* **2015**, *23*, 159–165. [CrossRef]
40. Manchikanti, L.; Singh, A. Therapeutic opioids: A ten-year perspective on the complexities and complications of the escalating use, abuse, and nonmedical use of opioids. *Pain Physician* **2008**, *11*, S63–S88. [CrossRef] [PubMed]
41. Duarte, R.V.; Raphael, J.H.; Southall, J.L.; Labib, M.H.; Whallett, A.J.; Ashford, R.L. Hypogonadism and low bone mineral density in patients on long-term intrathecal opioid delivery therapy. *BMJ Open* **2013**, *3*, e002856. [CrossRef] [PubMed]
42. Brennan, M.J. The Effect of Opioid Therapy on Endocrine Function. *Am. J. Med.* **2013**, *126*, S12–S18. [CrossRef] [PubMed]
43. Nelson, R.E.; Nebeker, J.R.; Sauer, B.; LaFleur, J. Factors associated with screening or treatment initiation among male United States veterans at risk for osteoporosis fracture. *Bone* **2012**, *50*, 983–988. [CrossRef] [PubMed]
44. Boshra, V. Evaluation of Osteoporosis Risk Associated with Chronic Use of Morphine, Fentanyl and Tramadol in Adult Female Rats. *Curr. Drug Saf.* **2011**, *6*, 159–163. [CrossRef] [PubMed]
45. Ramli, F.F.; Hashim, S.A.S.; Effendy, N.M. Factors Associated with Low Bone Density in Opioid Substitution Therapy Patients: A Systematic Review. *Int. J. Med Sci.* **2021**, *18*, 575–581. [CrossRef]
46. Vestergaard, P.; Rejnmark, L.; Mosekilde, L. Fracture risk associated with the use of morphine and opiates. *J. Intern. Med.* **2006**, *260*, 76–87. [CrossRef] [PubMed]
47. Coluzzi, F.; Scerpa, M.S.; Centanni, M. The Effect of Opiates on Bone Formation and Bone Healing. *Curr. Osteoporos. Rep.* **2020**, *18*, 325–335. [CrossRef] [PubMed]
48. Ezzatabadipour, M.; Majidi, M.; Malekpour-Afshar, R.; Eftekharvaghefi, S.H.; Nematollahi-Mahani, S.N. The Effects of Morphine on Tissue Structure of the Growth Plate in Male Rats. *Iran. J. Basic Med Sci.* **2011**, *14*, 514–520. [PubMed]
49. Saeidinezhad, M.; Razban, V.; Safizadeh, H.; Ezzatabadipour, M. Effects of maternal consumption of morphine on rat skeletal system development. *BMC Musculoskelet. Disord.* **2021**, *22*, 1–10. [CrossRef]

50. Kim, T.W.; Alford, D.P.; Malabanan, A.; Holick, M.; Samet, J. Low bone density in patients receiving methadone maintenance treatment. *Drug Alcohol Depend.* **2006**, *85*, 258–262. [CrossRef]
51. Kim, E.Y.; Kwon, D.H.; Lee, B.D.; Kim, Y.T.; Ahn, Y.B.; Yoon, K.Y.; Sa, S.J.; Cho, W.; Cho, S.N. Frequency of osteoporosis in 46 men with methamphetamine abuse hospitalized in a National Hospital. *Forensic Sci. Int.* **2009**, *188*, 75–80. [CrossRef] [PubMed]
52. Grey, A.; Rix-Trott, K.; Horne, A.; Gamble, G.; Bolland, M.; Reid, I. Decreased bone density in men on methadone maintenance therapy. *Addiction* **2010**, *106*, 349–354. [CrossRef]
53. Milos, G.; Gallo, L.M.; Sosic, B.; Uebelhart, D.; Goerres, G.; Haeuselmann, H.-J.; Eich, M. Bone Mineral Density in Young Women on Methadone Substitution. *Calcif. Tissue Res.* **2011**, *89*, 228–233. [CrossRef]
54. Pérez-Castrillón, J.L.; Olmos, J.M.; Gómez, J.J.; Barrallo-Gimeno, A.; Riancho, J.A.; Perera, L.; Valero, C.; Amado, J.A.; González-Macías, J. Expression of Opioid Receptors in Osteoblast-Like MG-63 Cells, and Effects of Different Opioid Agonists on Alkaline Phosphatase and Osteocalcin Secretion by These Cells. *Neuroendocrinology* **2000**, *72*, 187–194. [CrossRef]
55. Pedrazzoni, M.; Vescovi, P.P.; Maninetti, L.; Michelini, M.; Zaniboni, G.; Pioli, G.; Costi, D.; Alfano, F.S.; Passeri, M. Effects of chronic heroin abuse on bone and mineral metabolism. *Eur. J. Endocrinol.* **1993**, *129*, 42–45. [CrossRef]
56. Wilczek, H.; Stěpán, J. [Bone metabolism in individuals dependent on heroin and after methadone administration]. *Cas. Lek. Ceskych* **2003**, *142*, 606–608.
57. Brancós, M.A.; Peris, P.; Miró, J.; Monegal, A.; Gatell, J.; Mallolas, J.; Mensa, J.; García, S.; Muñoz-Gómez, J. Septic arthritis in heroin addicts. *Semin. Arthritis Rheum.* **1991**, *21*, 81–87. [CrossRef]
58. Allison, D.C.; Holtom, P.D.; Patzakis, M.J.; Zalavras, C.G. Microbiology of Bone and Joint Infections in Injecting Drug Abusers. *Clin. Orthop. Relat. Res.* **2010**, *468*, 2107–2112. [CrossRef] [PubMed]
59. Ross, J.J.; Hu, L.T. Septic arthritis of the pubic symphysis: Review of 100 cases. *Medicine* **2003**, *82*, 340–345. [CrossRef] [PubMed]
60. Delaney, F.T.; Stanley, E.; Bolster, F. The needle and the damage done: Musculoskeletal and vascular complications associated with injected drug use. *Insights Imaging* **2020**, *11*, 1–14. [CrossRef] [PubMed]
61. Endress, C.; Guyot, D.R.; Fata, J.; Salciccioli, G. Cervical osteomyelitis due to i.v. heroin use: Radiologic findings in 14 patients. *Am. J. Roentgenol.* **1990**, *155*, 333–335. [CrossRef]
62. Singh, G.; Shetty, R.R.; Ravidass, M.J.; Anilkumar, P.G. Cervical osteomyelitis associated with intravenous drug use. *Emerg. Med. J.* **2006**, *23*, e16. [CrossRef]
63. Grund, J.-P.C.; Latypov, A.; Harris, M. Breaking worse: The emergence of krokodil and excessive injuries among people who inject drugs in Eurasia. *Int. J. Drug Policy* **2013**, *24*, 265–274. [CrossRef]
64. Poghosyan, Y.M.; Hakobyan, K.A.; Poghosyan, A.Y.; Avetisyan, E.K. Surgical treatment of jaw osteonecrosis in "Krokodil" drug addicted patients. *J. Cranio-Maxillofac. Surg.* **2014**, *42*, 1639–1643. [CrossRef]
65. Hakobyan, K.; Poghosyan, Y. Spontaneous bone formation after mandible segmental resection in "krokodil" drug-related jaw osteonecrosis patient: Case report. *Oral Maxillofac. Surg.* **2017**, *21*, 267–270. [CrossRef]
66. Sergent, J.-F.; Bader, G.; Hamon, J.; Peigne, L.; Lejeune, S. Krokodil (Desomorphine)-induced osteonecrosis of the maxilla: A case report and literature review. *J. Oral Med. Oral Surg.* **2019**, *25*, 26. [CrossRef]
67. Tomita, M.; Katsuyama, H.; Watanabe, Y.; Okuyama, T.; Fushimi, S.; Ishikawa, T.; Nata, M.; Miyamoto, O. Does methamphetamine affect bone metabolism? *Toxicology* **2014**, *319*, 63–68. [CrossRef] [PubMed]
68. Katsuragawa, K. Effect of methamphetamine abuse on the bone quality of the calcaneus. *Forensic Sci. Int.* **1999**, *101*, 43–48. [CrossRef]
69. Mosti, M.P.; Flemmen, G.; Hoff, J.; Stunes, A.K.; Syversen, U.; Wang, E. Impaired skeletal health and neuromuscular function among amphetamine users in clinical treatment. *Osteoporos. Int.* **2016**, *27*, 1003–1010. [CrossRef] [PubMed]
70. Rustemeyer, J.; Melenberg, A.; Junker, K.; Sari-Rieger, A. Osteonecrosis of the maxilla related to long-standing methamphetamine abuse: A possible new aspect in the etiology of osteonecrosis of the jaw. *Oral Maxillofac. Surg.* **2014**, *18*, 237–241. [CrossRef]
71. Faucett, E.A.; Marsh, K.M.; Farshad, K.; Erman, A.B.; Chiu, A.G. Maxillary Sinus Manifestations of Methamphetamine Abuse. *Allergy Rhinol.* **2015**, *6*, 76–79. [CrossRef] [PubMed]
72. Ehrenkranz, J.; Levine, M.A. Bones and Joints: The Effects of Cannabinoids on the Skeleton. *J. Clin. Endocrinol. Metab.* **2019**, *104*, 4683–4694. [CrossRef]
73. Idris, A.I.; Ralston, S.H. Role of cannabinoids in the regulation of bone remodeling. *Front. Endocrinol.* **2012**, *3*, 136. [CrossRef]
74. Sophocleous, A.; Robertson, R.; Ferreira, N.B.; McKenzie, J.; Fraser, W.D.; Ralston, S.H. Heavy Cannabis Use Is Associated With Low Bone Mineral Density and an Increased Risk of Fractures. *Am. J. Med.* **2017**, *130*, 214–221. [CrossRef]
75. Bab, I.; Zimmer, A.; Melamed, E. Cannabinoids and the skeleton: From marijuana to reversal of bone loss. *Ann. Med.* **2009**, *41*, 560–567. [CrossRef]
76. Bourne, D.; Plinke, W.; Hooker, E.R.; Nielson, C.M. Cannabis use and bone mineral density: NHANES 2007–2010. *Arch. Osteoporos.* **2017**, *12*, 29. [CrossRef]
77. O'Connor, C.M.; Anoushiravani, A.A.; Adams, C.; Young, J.; Richardson, K.; Rosenbaum, A.J. Cannabinoid Use in Musculoskeletal Illness: A Review of the Current Evidence. *Curr. Rev. Musculoskelet. Med.* **2020**, *13*, 379–384. [CrossRef] [PubMed]
78. Sampson, H.W. Alcohol's harmful effects on bone. *Alcohol Health Res. World* **1998**, *22*, 190–194. [PubMed]
79. Maurel, D.B.; Boisseau, N.; Benhamou, C.L.; Jaffre, C. Alcohol and bone: Review of dose effects and mechanisms. *Osteoporos. Int.* **2012**, *23*, 1–16. [CrossRef] [PubMed]
80. Mikosch, P. Alcohol and bone. *Wien. Med. Wochenschr.* **2014**, *164*, 15–24. [CrossRef] [PubMed]

81. Klein, R.F. Alcohol-Induced Bone Disease: Impact of Ethanol on Osteoblast Proliferation. *Alcohol. Clin. Exp. Res.* **1997**, *21*, 392–399. [CrossRef] [PubMed]
82. Luo, Z.; Liu, Y.; Liu, Y.; Chen, H.; Shi, S.; Liu, Y. Cellular and molecular mechanisms of alcohol-induced osteopenia. *Cell. Mol. Life Sci.* **2017**, *74*, 4443–4453. [CrossRef]
83. Cheraghi, Z.; Doosti-Irani, A.; Almasi-Hashiani, A.; Baigi, V.; Mansournia, N.; Etminan, M.; Mansournia, M.A. The effect of alcohol on osteoporosis: A systematic review and meta-analysis. *Drug Alcohol Depend.* **2019**, *197*, 197–202. [CrossRef]
84. Gaddini, G.W.; Turner, R.T.; Grant, K.A.; Iwaniec, U.T. Alcohol: A Simple Nutrient with Complex Actions on Bone in the Adult Skeleton. *Alcohol. Clin. Exp. Res.* **2016**, *40*, 657–671. [CrossRef]
85. Jin, L.H.; Chang, S.J.; Koh, S.B.; Kim, K.S.; Lee, T.Y.; Ryu, S.Y.; Song, J.S.; Park, J.K. Association between alcohol consumption and bone strength in Korean adults: The Korean Genomic Rural Cohort Study. *Metabolism* **2011**, *60*, 351–358. [CrossRef]
86. Prieto-Alhambra, D.; Turkiewicz, A.; Reyes, C.; Timpka, S.; Rosengren, B.; Englund, M. Smoking and Alcohol Intake but Not Muscle Strength in Young Men Increase Fracture Risk at Middle Age: A Cohort Study Linked to the Swedish National Patient Registry. *J. Bone Miner. Res.* **2019**, *35*, 498–504. [CrossRef]
87. Richards, C.J.; Graf, K.W.; Mashru, R.P. The Effect of Opioids, Alcohol, and Nonsteroidal Anti-inflammatory Drugs on Fracture Union. *Orthop. Clin. North Am.* **2017**, *48*, 433–443. [CrossRef]
88. Jacobs, B. Alcoholism-induced bone necrosis. *N. Y. State J. Med.* **1992**, *92*, 334–338. [PubMed]
89. Matsumoto, K.; Ogawa, H.; Akiyama, H. Multifocal Osteonecrosis Secondary to Chronic Alcohol Ingestion. *Case Rep. Orthop.* **2015**, *2015*, 1–4. [CrossRef] [PubMed]
90. Richardson, S.; Browne, M.; Rasmussen, S.A.; Druschel, C.M.; Sun, L.; Jabs, E.; Romitti, P.A.; The National Birth Defects Prevention Study. Associations between periconceptional alcohol consumption and craniosynostosis, omphalocele, and gastroschisis. *Birth Defects Res. Part A Clin. Mol. Teratol.* **2011**, *91*, 623–630. [CrossRef]
91. Wong, P.K.K.; Christie, J.J.; Wark, J.D. The effects of smoking on bone health. *Clin. Sci.* **2007**, *113*, 233–241. [CrossRef] [PubMed]
92. Trevisan, C.; Alessi, A.; Girotti, G.; Zanforlini, B.M.; Bertocco, A.; Mazzochin, M.; Zoccarato, F.; Piovesan, F.; Dianin, M.; Giannini, S.; et al. The Impact of Smoking on Bone Metabolism, Bone Mineral Density and Vertebral Fractures in Postmenopausal Women. *J. Clin. Densitom.* **2020**, *23*, 381–389. [CrossRef]
93. Jaramillo, J.D.; Wilson, C.; Stinson, D.J.; Lynch, D.A.; Bowler, R.P.; Lutz, S.; Bon, J.M.; Arnold, B.; McDonald, M.-L.N.; Washko, G.R.; et al. Reduced Bone Density and Vertebral Fractures in Smokers. Men and COPD Patients at Increased Risk. *Ann. Am. Thorac. Soc.* **2015**, *12*, 648–656. [CrossRef] [PubMed]
94. Al-Bashaireh, A.M.; Haddad, L.G.; Weaver, M.; Chengguo, X.; Kelly, D.L.; Yoon, S. The Effect of Tobacco Smoking on Bone Mass: An Overview of Pathophysiologic Mechanisms. *J. Osteoporos.* **2018**, *2018*, 1–17. [CrossRef] [PubMed]
95. Strozyk, D.; Gress, T.M.; Breitling, L.P. Smoking and bone mineral density: Comprehensive analyses of the third National Health and Nutrition Examination Survey (NHANES III). *Arch. Osteoporos.* **2018**, *13*, 16. [CrossRef] [PubMed]
96. Mizrak, S.; Turan, V.; Inan, S.; Uysal, A.; Yilmaz, C.; Ercan, G. Effect of Nicotine on RANKL and OPG and Bone Mineral Density. *J. Investig. Surg.* **2014**, *27*, 327–331. [CrossRef] [PubMed]
97. Yuan, S.; Michaëlsson, K.; Wan, Z.; Larsson, S.C. Associations of Smoking and Alcohol and Coffee Intake with Fracture and Bone Mineral Density: A Mendelian Randomization Study. *Calcif. Tissue Res.* **2019**, *105*, 582–588. [CrossRef] [PubMed]
98. Ma, L.; Sham, M.H.; Zheng, L.W.; Cheung, L.K. Influence of Low-Dose Nicotine on Bone Healing. *J. Trauma Inj. Infect. Crit. Care* **2011**, *70*, E117–E121. [CrossRef] [PubMed]
99. Rodriguez-Merchan, E.C. The importance of smoking in orthopedic surgery. *Hosp. Pr.* **2018**, *46*, 175–182. [CrossRef] [PubMed]
100. Hernigou, J.; Schuind, F. Tobacco and bone fractures: A review of the facts and issues that every orthopaedic surgeon should know. *Bone Jt. Res.* **2019**, *8*, 255–265. [CrossRef] [PubMed]
101. Lorentzon, M.; Mellström, D.; Haug, E.; Ohlsson, C. Smoking is associated with lower bone mineral density and reduced cortical thickness in young men. *J. Clin. Endocrinol. Metab.* **2007**, *92*, 497–503. [CrossRef]
102. Scolaro, J.A.; Schenker, M.L.; Yannascoli, S.; Baldwin, K.; Mehta, S.; Ahn, J. Cigarette smoking increases complications following fracture: A systematic review. *J. Bone Jt. Surg. Am. Vol.* **2014**, *96*, 674–681. [CrossRef] [PubMed]
103. Barbosa, A.P.; Lourenço, J.D.; Junqueira, J.J.M.; de França, S.L.E.; Martins, J.S.; Junior, M.C.O.; Begalli, I.; Velosa, A.P.P.; Olivo, C.R.; Bastos, T.B.; et al. The deleterious effects of smoking in bone mineralization and fibrillar matrix composition. *Life Sci.* **2020**, *241*, 117132. [CrossRef]
104. Aspera-Werz, R.H.; Chen, T.; Ehnert, S.; Zhu, S.; Fröhlich, T.; Nussler, A.K. Cigarette Smoke Induces the Risk of Metabolic Bone Diseases: Transforming Growth Factor Beta Signaling Impairment via Dysfunctional Primary Cilia Affects Migration, Proliferation, and Differentiation of Human Mesenchymal Stem Cells. *Int. J. Mol. Sci.* **2019**, *20*, 2915. [CrossRef] [PubMed]
105. Ding, C.; Cicuttini, F.; Blizzard, L.; Jones, G. Smoking interacts with family history with regard to change in knee cartilage volume and cartilage defect development. *Arthritis Care Res.* **2007**, *56*, 1521–1528. [CrossRef] [PubMed]
106. Amin, S.; Niu, J.; Guermazi, A.; Grigoryan, M.; Hunter, D.J.; Clancy, M.; LaValley, M.; Genant, H.K.; Felson, D. Cigarette smoking and the risk for cartilage loss and knee pain in men with knee osteoarthritis. *Ann. Rheum. Dis.* **2006**, *66*, 18–22. [CrossRef]
107. Wang, D.; Nasto, L.; Roughley, P.; Leme, A.; Houghton, A.; Usas, A.; Sowa, G.; Lee, J.; Niedernhofer, L.; Shapiro, S.; et al. Spine degeneration in a murine model of chronic human tobacco smokers. *Osteoarthr. Cartil.* **2012**, *20*, 896–905. [CrossRef]

108. Quelhas, D.; Kompala, C.; Wittenbrink, B.; Han, Z.; Parker, M.; Shapiro, M.; Downs, S.; Kraemer, K.; Fanzo, J.; Morris, S.; et al. The association between active tobacco use during pregnancy and growth outcomes of children under five years of age: A systematic review and meta-analysis. *BMC Public Heal.* **2018**, *18*, 1372. [CrossRef]
109. Bowles, S. *Drug Induced Osteoporosis*; Pharmacotherapy Self-Assessment Program, American College of Clinical Pharmacy: Kansas City, MO, USA, 2012; pp. 203–224.
110. Spoelhof, B.; Ray, S.D. Corticosteroids. In *Encyclopedia of Toxicology: Third Edition*; Wexler, P., Ed.; Academic Press: London, UK, 2014; pp. 1038–1042.
111. Park, H.W.; Dahlin, A.; Tse, S.; Daun, Q.L.; Schuemann, B.; Martinez, F.D.; Peters, S.P.; Szefler, S.J.; Lima, J.J.; Kubo, M.; et al. Genetic procedures associated with improvement of asthma symptoms in response to inhaled corticosteroids. *J. Allergy Clin. Immunol.* **2014**, *133*, 664–669. [CrossRef] [PubMed]
112. Romas, E. Corticosteroid-induced osteoporosis and fractures. *Aust. Prescr.* **2008**, *31*, 45–49. [CrossRef]
113. Liu, Y.-Z.; Akhter, M.P.; Gao, X.; Wang, X.-Y.; Zhao, G.; Wei, X.; Wu, H.-J.; Chen, H.; Wang, D.; Cui, L. Glucocorticoid-induced delayed fracture healing and impaired bone biomechanical properties in mice. *Clin. Interv. Aging* **2018**, *13*, 1465–1474. [CrossRef] [PubMed]
114. Allen, D.B.; Julius, J.R.; Breen, T.J.; Attie, K.M. Treatment of glucocorticoid-induced growth suppression with growth hormone. *J. Clin. Endocrinol. Metab.* **1998**, *83*, 2824–2829. [CrossRef] [PubMed]
115. Annexstad, E.J.; Bollerslev, J.; Westvik, J.; Myhre, A.G.; Godang, K.; Holm, I.; Rasmussen, M. The role of delayed bone age in the evaluation of stature and bone health in glucocorticoid treated patients with Duchenne muscular dystrophy. *Int. J. Pediatr. Endocrinol.* **2019**, *2019*, 1–12. [CrossRef] [PubMed]
116. Mushtaq, T.; Ahmed, S.F. The impact of corticosteroids on growth and bone health. *Arch. Dis. Child.* **2002**, *87*, 93–96. [CrossRef] [PubMed]
117. Wang, T.; Li, Y.; Ye, Y.-Y.; Huang, H.; Yi, H.-L.; Chen, M.; Guo, C. Effects of inhaled corticosteroids on bone age and growth in children with asthma. *Chin. J. Contemp. Pediatrics* **2012**, *14*, 359–361. (In Chinese)
118. Black, S.; Aggrawal, A.; Payne-James, J. (Eds.) *Age Estimation in the Living: A Practitioner's Guide*; Wiley-Blackwell: Chichester, UK, 2012.
119. García-Martínez, O.; De Luna-Bertos, E.; Ramos-Torrecillas, J.; Manzano-Moreno, F.; Ruiz, C. Repercussions of NSAIDS drugs on bone tissue: The osteoblast. *Life Sci.* **2015**, *123*, 72–77. [CrossRef] [PubMed]
120. Nwadinigwe, C.U.; Anyaehie, U.E. Effects of Cyclooxygenase inhibitors on bone and cartilage metabolism: A Review. *Niger. J. Med.* **2008**, *16*, 290–294. [CrossRef]
121. Lisowska, B.; Kosson, D.; Domaracka, K. Lights and shadows of NSAIDs in bone healing: The role of prostaglandins in bone metabolism. *Drug Des. Dev. Ther.* **2018**, *12*, 1753–1758. [CrossRef] [PubMed]
122. Lisowska, B.; Kosson, D.; Domaracka, K. Positives and negatives of nonsteroidal anti-inflammatory drugs in bone healing: The effects of these drugs on bone repair. *Drug Des. Dev. Ther.* **2018**, *12*, 1809–1814. [CrossRef] [PubMed]
123. O'Connor, J.P.; Lysz, T. Celecoxib, NSAIDs and the skeleton. *Drugs Today* **2008**, *44*, 693–709. [CrossRef] [PubMed]
124. Dodwell, E.R.; Latorre, J.G.; Parisini, E.; Zwettler, E.; Chandra, D.; Mulpuri, K.; Snyder, B. NSAID Exposure and Risk of Nonunion: A Meta-Analysis of Case–Control and Cohort Studies. *Calcif. Tissue Res.* **2010**, *87*, 193–202. [CrossRef] [PubMed]
125. Goodman, S.B.; Ma, T.; Genovese, M.; Smith, R.L. Cox-2 Selective Inhibitors and Bone. *Int. J. Immunopathol. Pharmacol.* **2003**, *16*, 201–205. [CrossRef] [PubMed]
126. Geusens, P.; Emans, P.J.; de Jong, J.J.; van den Bergh, J.A. NSAIDs and fracture healing. *Curr. Opin. Rheumatol.* **2013**, *25*, 524–531. [CrossRef] [PubMed]
127. Pountos, I.; Georgouli, T.; Calori, G.M.; Giannoudis, P.V. Do Nonsteroidal Anti-Inflammatory Drugs Affect Bone Healing? A Critical Analysis. *Sci. World J.* **2012**, *2012*, 1–14. [CrossRef] [PubMed]
128. Vuolteenaho, K.; Moilanen, T.; Moilanen, E. Non-Steroidal Anti-Inflammatory Drugs, Cyclooxygenase-2 and the Bone Healing Process. *Basic Clin. Pharmacol. Toxicol.* **2008**, *102*, 10–14. [CrossRef] [PubMed]
129. Wheatley, B.M.; Nappo, K.E.; Christensen, D.L.; Holman, A.M.; Brooks, D.I.; Potter, B.K. Effect of NSAIDs on bone healing rates: A meta-analysis. *J. Am. Acad. Orthop. Surg.* **2019**, *27*, e330–e336. [CrossRef] [PubMed]
130. Chen, M.R.; Dragoo, J.L. The effect of nonsteroidal anti-inflammatory drugs on tissue healing. *Knee Surgery, Sports Traumatol. Arthrosc.* **2012**, *21*, 540–549. [CrossRef] [PubMed]
131. Bhattacharyya, T.; Levin, R.; Vrahas, M.S.; Solomon, D.H. Nonsteroidal antiinflammatory drugs and nonunion of humeral shaft fractures. *Arthritis Care Res.* **2005**, *53*, 364–367. [CrossRef] [PubMed]
132. Jeffcoach, D.R.; Sams, V.G.; Lawson, C.M.; Enderson, B.L.; Smith, S.T.; Kline, H.; Barlow, P.B.; Wylie, D.R.; Krumenacker, L.A.; McMillen, J.C.; et al. Nonsteroidal anti-inflammatory drugs' impact on nonunion and infection rates in long-bone fractures. *J. Trauma Acute Care Surg.* **2014**, *76*, 779–783. [CrossRef] [PubMed]
133. Marquez-Lara, A.; Hutchinson, I.D.; Nuñez, F.; Smith, T.L.; Miller, A.N. Nonsteroidal Anti-Inflammatory Drugs and Bone-Healing. *JBJS Rev.* **2016**, *4*, e4. [CrossRef]
134. Salari, P.; Abdollahi, M. Controversial effects of non-steroidal anti-inflammatory drugs on bone: A review. *Inflamm. Allergy-Drug Targets* **2009**, *8*, 169–175. [CrossRef] [PubMed]
135. Borgeat, A.; Ofner, C.; Saporito, A.; Farshad, M.; Aguirre, J. The effect of nonsteroidal anti-inflammatory drugs on bone healing in humans: A qualitative, systematic review. *J. Clin. Anesthesia* **2018**, *49*, 92–100. [CrossRef]

136. Xie, Y.; Pan, M.; Gao, Y.; Zhang, L.; Ge, W.; Tang, P. Dose-dependent roles of aspirin and other non-steroidal anti-inflammatory drugs in abnormal bone remodeling and skeletal regeneration. *Cell Biosci.* **2019**, *9*, 1–11. [CrossRef]
137. Bjelle, A. NSAIDs and Cartilage Metabolism. *Scand. J. Rheumatol.* **1988**, *18*, 43–52. [CrossRef]
138. Ding, C.; Cicuttini, F.; Jones, G. Do NSAIDs Affect Longitudinal Changes in Knee Cartilage Volume and Knee Cartilage Defects in Older Adults? *Am. J. Med.* **2009**, *122*, 836–842. [CrossRef]
139. Vestergaard, P.; Rejnmark, L.; Mosekilde, L. Fracture Risk Associated with Use of Nonsteroidal Anti-Inflammatory Drugs, Acetylsalicylic Acid, and Acetaminophen and the Effects of Rheumatoid Arthritis and Osteoarthritis. *Calcif. Tissue Res.* **2006**, *79*, 84–94. [CrossRef]
140. Vestergaard, P. Pain-relief medication and risk of fractures. *Curr. Drug Saf.* **2008**, *3*, 199–203. [CrossRef] [PubMed]
141. Aliuskevicius, M.; Østgaard, S.E.; Hauge, E.M.; Vestergaard, P.; Rasmussen, S. Influence of Ibuprofen on Bone Healing After Colles' Fracture: A Randomized Controlled Clinical Trial. *J. Orthop. Res.* **2020**, *38*, 545–554. [CrossRef] [PubMed]
142. Konstantinidis, I.; Papageorgiou, S.N.; Kyrgidis, A.; Tzellos, T.-G.; Kouvelas, D. Effect of non-steroidal anti-inflammatory drugs on bone turnover: An evidence-based review. *Rev. Recent Clin. Trials* **2013**, *8*, 48–60. [CrossRef] [PubMed]
143. Vestergaard, P.; Hermann, P.; Jensen, J.-E.B.; Eiken, P.; Mosekilde, L. Effects of paracetamol, non-steroidal anti-inflammatory drugs, acetylsalicylic acid, and opioids on bone mineral density and risk of fracture: Results of the Danish Osteoporosis Prevention Study (DOPS). *Osteoporos. Int.* **2011**, *23*, 1255–1265. [CrossRef] [PubMed]
144. Chin, K.-Y. A Review on the Relationship between Aspirin and Bone Health. *J. Osteoporos.* **2017**, *2017*, 1–8. [CrossRef] [PubMed]
145. Nuelle, J.A.; Coe, K.M.; Oliver, H.A.; Cook, J.L.; Hoernschemeyer, D.G.; Gupta, S.K. Effect of NSAID Use on Bone Healing in Pediatric Fractures: A Preliminary, Prospective, Randomized, Blinded Study. *J. Pediatr. Orthop.* **2020**, *40*, e683–e689. [CrossRef]
146. Antonucci, R.; Zaffanello, M.; Puxeddu, E.; Porcella, A.; Cuzzolin, L.; Pilloni, M.D.; Fanos, V. Use of Non-steroidal Anti-inflammatory Drugs in Pregnancy: Impact on the Fetus and Newborn. *Curr. Drug Metab.* **2012**, *13*, 474–490. [CrossRef]
147. Williams, L.J.; Pasco, J.A.; Henry, M.J.; Sanders, K.M.; Nicholson, G.C.; Kotowicz, M.A.; Berk, M. Paracetamol (acetaminophen) use, fracture and bone mineral density. *Bone* **2011**, *48*, 1277–1281. [CrossRef]
148. Díaz-Rodríguez, L.; García-Martínez, O.; Arroyo-Morales, M.; Rubio-Ruiz, B.; Ruiz, C. Effect of acetaminophen (paracetamol) on human osteosarcoma cell line MG. *Acta Pharmacol. Sin.* **2010**, *31*, 1495–1499. [CrossRef]
149. Richards, J.B.; Joseph, L.; Schwartzman, K.; Kreiger, N.; Tenenhouse, A.; Goltzman, D. The effect of cyclooxygenase-2 inhibitors on bone mineral density: Results from the Canadian Multicentre Osteoporosis Study. *Osteoporos. Int.* **2006**, *17*, 1410–1419. [CrossRef]
150. Yoshida, K.; Yu, Z.; Greendale, G.A.; Ruppert, K.; Lian, Y.; Tedeschi, S.K.; Lin, T.-C.; Haneuse, S.; Glynn, R.J.; Hernández-Díaz, S.; et al. Effects of analgesics on bone mineral density: A longitudinal analysis of the prospective SWAN cohort with three-group matching weights. *Pharmacoepidemiol. Drug Saf.* **2018**, *27*, 182–190. [CrossRef] [PubMed]
151. Faienza, M.F.; Brunetti, G.; Acquafredda, A.; Delvecchio, M.; Lonero, A.; Gaeta, A.; Bulzis, P.S.; Corica, D.; Velletri, M.R.; De Luca, F.; et al. Metabolic Outcomes, Bone Health, and Risk of Polycystic Ovary Syndrome in Girls with Idiopathic Central Precocious Puberty Treated with Gonadotropin-Releasing Hormone Analogues. *Horm. Res. Paediatr.* **2017**, *87*, 162–169. [CrossRef] [PubMed]
152. Rachner, T.D.; Coleman, R.; Hadji, P.; Hofbauer, L.C. Bone health during endocrine therapy for cancer The Lancet. *Diabetes Endocrinol.* **2018**, *6*, 901–910.
153. Mohamad, N.-V.; Ima-Nirwana, S.; Chin, K.-Y. The effects of gonadotropin-releasing hormone agonist (buserelin) and orchidectomy on bone turnover markers and histomorphometry in rats. *Aging Male* **2020**, *23*, 327–334. [CrossRef] [PubMed]
154. Handforth, C.; D'Oronzo, S.; Coleman, R.; Brown, J. Cancer Treatment and Bone Health. *Calcif. Tissue Res.* **2018**, *102*, 251–264. [CrossRef] [PubMed]
155. Sauerbrun-Cutler, M.-T.; Alvero, R. Short- and long-term impact of gonadotropin-releasing hormone analogue treatment on bone loss and fracture. *Fertil. Steril.* **2019**, *112*, 799–803. [CrossRef]
156. Shapiro, C.L. Osteoporosis: A Long-Term and Late-Effect of Breast Cancer Treatments. *Cancers* **2020**, *12*, 3094. [CrossRef]
157. Smith, M.R. Therapy Insight: Osteoporosis during hormone therapy for prostate cancer. *Nat. Clin. Pr. Urol.* **2005**, *2*, 608–615. [CrossRef]
158. Wallach, E.E.; Fogelman, I. Gonadotropin-releasing hormone agonists and the skeleton. *Fertil. Steril.* **1992**, *57*, 715–724. [CrossRef]
159. Johansen, J.S.; Riis, B.J.; Hassager, C.; Moen, M.; Jacobson, J.; Christiansen, C. The Effect of a Gonadotropin-Releasing Hormone Agonist Analog (Nafarelin) on Bone Metabolism. *J. Clin. Endocrinol. Metab.* **1988**, *67*, 701–706. [CrossRef]
160. Nencioni, T.; Penotti, M.; Barbieri-Carones, M.; Ortolani, S.; Trevisan, C.; Polvani, F. Gonadotropin releasing hormone agonist therapy and its effect on bone mass. *Gynecol. Endocrinol.* **1991**, *5*, 49–56. [CrossRef] [PubMed]
161. Uemura, T.; Mohri, J.; Osada, H.; Suzuki, N.; Katagiri, N.; Minaguchi, H. Effect of gonadotropin-releasing hormone agonist on the bone mineral density of patients with endometriosis*†. *Fertil. Steril.* **1994**, *62*, 246–250. [CrossRef]
162. Antoniazzi, F.; Monti, E.; Gaudino, R.; Cavarzere, P.; Zaffanello, M.; Brugnara, M.; Perlini, S.; Maines, E.; Gallo, M.C.; Corso, S.D.; et al. Bone density in children treated with gonadotropin-releasing hormone analogs for central precocious puberty. *Expert Rev. Endocrinol. Metab.* **2010**, *5*, 285–290. [CrossRef] [PubMed]
163. Stevenson, M.O.; Tangpricha, V. Osteoporosis and Bone Health in Transgender Persons. *Endocrinol. Metab. Clin. North Am.* **2019**, *48*, 421–427. [CrossRef] [PubMed]
164. Klink, D.; Caris, M.; Heijboer, A.; Van Trotsenburg, M.; Rotteveel, J. Bone Mass in Young Adulthood Following Gonadotropin-Releasing Hormone Analog Treatment and Cross-Sex Hormone Treatment in Adolescents With Gender Dysphoria. *J. Clin. Endocrinol. Metab.* **2015**, *100*, E270–E275. [CrossRef] [PubMed]

165. Saggese, G.; Bertelloni, S.; Baroncelli, G.I.; Battini, R.; Franchi, G. Reduction of bone density: An effect of gonadotropin releasing hormone analogue treatment in central precocious puberty. *Eur. J. Pediatr.* **1993**, *152*, 717–720. [CrossRef]
166. Liu, J.; Li, X.; Fan, L.; Yang, J.; Wang, J.; Sun, J.; Wang, Z. Proton pump inhibitors therapy and risk of bone diseases: An update meta-analysis. *Life Sci.* **2019**, *218*, 213–223. [CrossRef]
167. Mattsson, J.P.; Väänänen, K.; Wallmark, B.; Lorentzon, P. Omeprazole and bafilomycin, two proton pump inhibitors: Differentiation of their effects on gastric, kidney and bone H+-translocating ATPases. *Biochim. et Biophys. Acta (BBA) Biomembr.* **1991**, *1065*, 261–268. [CrossRef]
168. Leontiadis, G.I.; Moayyedi, P. Proton Pump Inhibitors and Risk of Bone Fractures. *Curr. Treat. Options Gastroenterol.* **2014**, *12*, 414–423. [CrossRef]
169. Arj, A.; Zade, M.R.; Yavari, M.; Akbari, H.; Zamani, B.; Asemi, Z. Proton pump inhibitors use and change in bone mineral density. *Int. J. Rheum. Dis.* **2016**, *19*, 864–868. [CrossRef]
170. Targownik, L.E.; Goertzen, A.L.; Luo, Y.; Leslie, W. Long-Term Proton Pump Inhibitor Use Is Not Associated With Changes in Bone Strength and Structure. *Am. J. Gastroenterol.* **2017**, *112*, 95–101. [CrossRef] [PubMed]
171. Samji, H.; Cescon, A.; Hogg, R.S.; Modur, S.P.; Althoff, K.; Buchacz, K.; Burchell, A.N.; Cohen, M.; Gebo, K.A.; Gill, M.J.; et al. Closing the Gap: Increases in Life Expectancy among Treated HIV-Positive Individuals in the United States and Canada. *PLoS ONE* **2013**, *8*, e81355. [CrossRef]
172. Biehler-Gomez, L.; Cabrini, A.; De Angelis, D.; Cattaneo, C. How do skeletons with HIV present? A study on the identified CAL Milano Cemetery Skeletal Collection. *Leg. Med.* **2018**, *33*, 11–16. [CrossRef]
173. Cotter, A.G.; Powderly, W.G. Endocrine complications of human immunodeficiency virus infection: Hypogonadism, bone disease and tenofovir-related toxicity. *Best Pr. Res. Clin. Endocrinol. Metab.* **2011**, *25*, 501–515. [CrossRef] [PubMed]
174. Powderly, W.G. Osteoporosis and Bone Health in HIV. *Curr. HIV/AIDS Rep.* **2012**, *9*, 218–222. [CrossRef]
175. Molia, A.C.; Strady, C.; Rouger, C.; Beguinot, I.M.; Berger, J.-L.; Trenque, T.C. Osteonecrosis in Six HIV-Infected Patients Receiving Highly Active Antiretroviral Therapy. *Ann. Pharmacother.* **2004**, *38*, 2050–2054. [CrossRef] [PubMed]
176. Borges, Á.H.; Hoy, J.; Florence, E.; Sedlacek, D.; Stellbrink, H.-J.; Uzdaviniene, V.; Tomazic, J.; Gargalianos-Kakolyris, P.; Schmid, P.; Orkin, C.; et al. Antiretrovirals, Fractures, and Osteonecrosis in a Large International HIV Cohort. *Clin. Infect. Dis.* **2017**, *64*, 1413–1421. [CrossRef]
177. Mccomsey, G.A.; Tebas, P.; Shane, E.; Yin, M.T.; Overton, E.T.; Huang, J.; Aldrovandi, G.M.; Cardoso, S.W.; Santana, J.L.; Brown, T.T. Bone Disease in HIV Infection: A Practical Review and Recommendations for HIV Care Providers. *Clin. Infect. Dis.* **2010**, *51*, 937–946. [CrossRef] [PubMed]
178. Güerri-Fernández, R.; Villar-García, J.; Díez-Pérez, A.; Prieto-Alhambra, D. HIV infection, bone metabolism, and fractures. *Arq. Bras. Endocrinol. E Metabol.* **2014**, *58*, 478–483. [CrossRef]
179. Ahmad, A.N.; Ahmad, S.N.; Ahmad, N. HIV Infection and Bone Abnormalities. *Open Orthop. J.* **2017**, *11*, 777–784. [CrossRef]
180. Knobel, H.; Guelar, A.; Vallecillo, G.; Nogues, X.; Díez, A. Osteopenia in HIV-infected patients: Is it the disease or is it the treatment? *AIDS* **2001**, *15*, 807–808. [CrossRef]
181. Cazanave, C.; Dupon, M.; Lavignolle-Aurillac, V.; Barthe, N.; Lawson-Ayayi, S.; Mehsen, N.; Mercié, P.; Morlat, P.; Thiébaut, R.; Dabis, F. Reduced bone mineral density in HIV-infected patients: Prevalence and associated factors. *AIDS* **2008**, *22*, 395–402. [CrossRef] [PubMed]
182. Stone, B.; Dockrell, D.; Bowman, C.; McCloskey, E. HIV and bone disease. *Arch. Biochem. Biophys.* **2010**, *503*, 66–77. [CrossRef]
183. Ofotokun, I.; Weitzmann, M.N. HIV and bone metabolism. *Discov. Med.* **2011**, *11*, 385–393. [PubMed]
184. Greenberg, S.B.; Lahart, C.J. HIV/AIDS: Clinical Considerations. In *Kelley's Textbook of Internal Medicine*, 4th ed.; Humes, H.D., DuPont, H.L., Gardner, L.B., Griffin, J.W., Harris, E.D., Hazzard, W.R., King, T.E., Loriaux, D.L., et al., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2000; pp. 2320–2328.
185. Winias, S.; Radithia, D.; Ernawati, D.S. Neuropathy complication of antiretroviral therapy in HIV/AIDS patients. *Oral Dis.* **2020**, *26*, 149–152. [CrossRef] [PubMed]
186. Young, N.; Neiderer, K.; Martin, B.; Jolley, D.; Dancho, J.F. HIV neuropathy induced Charcot neuroarthropathy: A case discussion. *Foot* **2012**, *22*, 112–116. [CrossRef] [PubMed]
187. Rogers, L.C.; Bevilacqua, N.J.; Dellacorte, M.P.; Francis, K.; Armstrong, D.G. Charcot's arthropathy in a patient with HIV-associated neuropathy. *J. Am. Podiatr. Med. Assoc.* **2008**, *98*, 153–155. [PubMed]
188. Brody, D.J.; Gu, Q. *Antidepressant Use among adults: United States, 2015–NCHS Data Brief*; No National Center for Health Statistics: Hyattsville, MD, USA, 2020.
189. Bonnet, N.; Bernard, P.; Beaupied, H.; Bizot, J.C.; Trovero, F.; Courteix, D.; Benhamou, C.L. Various effects of antidepressant drugs on bone microarchitecture, mechanical properties and bone remodeling. *Toxicol. Appl. Pharmacol.* **2007**, *221*, 111–118. [CrossRef] [PubMed]
190. Williams, L.J.; Berk, M.; Hodge, J.M.; Kotowicz, M.A.; Stuart, A.L.; Chandrasekaran, V.; Cleminson, J.; Pasco, J.A. Selective Serotonin Reuptake Inhibitors (SSRIs) and Markers of Bone Turnover in Men. *Calcif. Tissue Res.* **2018**, *103*, 125–130. [CrossRef] [PubMed]
191. Tsapakis, E.; Gamie, Z.; Tran, G.T.; Adshead, S.; Lampard, A.; Mantalaris, A.; Tsiridis, E. The adverse skeletal effects of selective serotonin reuptake inhibitors. *Eur. Psychiatry* **2012**, *27*, 156–169. [CrossRef] [PubMed]

192. Rauma, P.H.; Pasco, J.A.; Berk, M.; Stuart, A.L.; Koivumaa-Honkanen, H.; Honkanen, R.J.; Hodge, J.M.; Williams, L.J. The association between use of antidepressants and bone quality using quantitative heel ultrasound. *Aust. N. Z. J. Psychiatry* **2015**, *49*, 437–443. [CrossRef] [PubMed]
193. Wadhwa, R.; Kumar, M.; Talegaonkar, S.; Vohora, D. Serotonin reuptake inhibitors and bone health: A review of clinical studies and plausible mechanisms. *Osteoporos. Sarcopenia* **2017**, *3*, 75–81. [CrossRef] [PubMed]
194. Bruyere, O.; Reginster, J.-Y. Osteoporosis in patients taking selective serotonin reuptake inhibitors: A focus on fracture outcome. *Endocrine* **2015**, *48*, 65–68. [CrossRef] [PubMed]
195. Rizzoli, R.; Cooper, C.; Reginster, J.-Y.; Abrahamsen, B.; Adachi, J.; Brandi, M.; Bruyere, O.; Compston, J.; Ducy, P.; Ferrari, S.; et al. Antidepressant medications and osteoporosis. *Bone* **2012**, *51*, 606–613. [CrossRef]
196. Sansone, R.A.; A Sansone, L. SSRIs: Bad to the bone? *Innov. Clin. Neurosci.* **2012**, *9*, 42–47. [PubMed]
197. Howie, R.N.; Herberg, S.; Durham, E.; Grey, Z.; Bennfors, G.; Elsalanty, M.; LaRue, A.C.; Hill, W.D.; Cray, J.J. Selective serotonin re-uptake inhibitor sertraline inhibits bone healing in a calvarial defect model. *Int. J. Oral Sci.* **2018**, *10*, 1–11. [CrossRef]
198. Abu Nada, L.; Al Subaie, A.; Mansour, A.; Wu, X.; Abdallah, M.-N.; Al-Waeli, H.; Ersheidat, A.; Stone, L.S.; Murshed, M.; Tamimi, F. The antidepressant drug, sertraline, hinders bone healing and osseointegration in rats' tibiae. *J. Clin. Periodontol.* **2018**, *45*, 1485–1497. [CrossRef]
199. Kurmanji, J.; Sulaiman, S.S.; Chandrasekaran, P.; Kah, L. PMH8 EFFECT OF VARIOUS ANTIDEPRESSANT GROUPS ON BONE MINERAL DENSITY (BMD). *Value Heal.* **2011**, *14*, A186. [CrossRef]
200. Gebara, M.A.; Shea, M.L.; Lipsey, K.L.; Teitelbaum, S.L.; Civitelli, R.; Müller, D.J.; Reynolds, C.F., 3rd; Mulsant, B.H.; Lenze, E.J. Depression, antidepressants, and bone health in older adults: A systematic review. *J. Am. Geriatr. Soc.* **2014**, *62*, 1434–1441. [CrossRef] [PubMed]
201. Chung, S.; Ahn, C. Effects of anti-epileptic drug therapy on bone mineral density in ambulatory epileptic children. *Brain Dev.* **1994**, *16*, 382–385. [CrossRef]
202. Gissel, T.; Poulsen, C.S.; Vestergaard, P. Adverse effects of antiepileptic drugs on bone mineral density in children. *Expert Opin. Drug Saf.* **2007**, *6*, 267–278. [CrossRef] [PubMed]
203. Zhang, Y.; Zheng, Y.-X.; Zhu, J.-M.; Zhang, J.-M.; Zheng, Z. Effects of antiepileptic drugs on bone mineral density and bone metabolism in children: A meta-analysis. *J. Zhejiang Univ. Sci. B* **2015**, *16*, 611–621. [CrossRef]
204. Zhong, R.; Chen, Q.; Zhang, X.; Li, M.; Liang, J.; Lin, W. Bone Mineral Density Loss in People With Epilepsy Taking Valproate as a Monotherapy: A Systematic Review and Meta-Analysis. *Front. Neurol.* **2019**, *10*. [CrossRef]
205. Sakellarides, M.; Bright, T.; Todaro, M.; Roten, A.; Lauren, J.D.; O'Brien, T.J.; John, D.W. Short-term vs. long-term duration of AED (anti-epileptic drug) pharmacotherapy: Effects on bone health parameters. *J. Clin. Neurosci.* **2009**, *16*, 1532–1533. [CrossRef]
206. Pack, A.M.; Morrell, M.J. Epilepsy and bone health in adults. *Epilepsy Behav.* **2004**, *5*, 24–29. [CrossRef] [PubMed]
207. Beerhorst, K.; Tan, I.Y.; De Krom, M.; Verschuure, P.; Aldenkamp, A.P. Antiepileptic drugs and high prevalence of low bone mineral density in a group of inpatients with chronic epilepsy. *Acta Neurol. Scand.* **2013**, *128*, 273–280. [CrossRef] [PubMed]
208. Khanna, S.; Pillai, K.K.; Vohora, D. Insights into liaison between antiepileptic drugs and bone. *Drug Discov. Today* **2009**, *14*, 428–435. [CrossRef]
209. Mattson, R.H.; Gidal, B.E. Fractures, epilepsy, and antiepileptic drugs. *Epilepsy Behav.* **2004**, *5*, 36–40. [CrossRef]
210. Nakken, K.; Taubøll, E. Bone loss associated with use of antiepileptic drugs. *Expert Opin. Drug Saf.* **2010**, *9*, 561–571. [CrossRef]
211. Pack, A.M. The association between antiepileptic drugs and bone disease. *Epilepsy Curr.* **2003**, *3*, 91–95. [CrossRef]
212. Gupta, Y.K.; Arora, E.; Singh, H. Impact of antiepileptic drugs on bone health: Need for monitoring, treatment, and prevention strategies. *J. Fam. Med. Prim. Care* **2016**, *5*, 248–253. [CrossRef] [PubMed]
213. Lee, H.-S.; Wang, S.-Y.; Salter, D.M.; Wang, C.-C.; Chen, S.-J.; Fan, H.-C. The impact of the use of antiepileptic drugs on the growth of children. *BMC Pediatr.* **2013**, *13*, 211. [CrossRef] [PubMed]
214. Elamir, Y.; Gianakos, A.L.; Lane, J.M.; Sharma, A.; Grist, W.P.; Liporace, F.A.; Yoon, R.S. The Effects of Diabetes and Diabetic Medications on Bone Health. *J. Orthop. Trauma* **2020**, *34*, e102–e108. [CrossRef] [PubMed]
215. Meier, C.; Schwartz, A.V.; Egger, A.; Lecka-Czernik, B. Effects of diabetes drugs on the skeleton. *Bone* **2016**, *82*, 93–100. [CrossRef]
216. Cortet, B.; Lucas, S.; Legroux-Gerot, I.; Penel, G.; Chauveau, C.; Paccou, J. Bone disorders associated with diabetes mellitus and its treatments. *Jt. Bone Spine* **2019**, *86*, 315–320. [CrossRef]
217. Gilbert, M.P.; Pratley, R.E. The Impact of Diabetes and Diabetes Medications on Bone Health. *Endocr. Rev.* **2015**, *36*, 194–213. [CrossRef]
218. Chandran, M. Diabetes Drug Effects on the Skeleton. *Calcif. Tissue Res.* **2016**, *100*, 133–149. [CrossRef]
219. Adil, M.; Khan, R.A.; Kalam, A.; Venkata, S.K.; Kandhare, A.; Ghosh, P.; Sharma, M. Effect of anti-diabetic drugs on bone metabolism: Evidence from preclinical and clinical studies. *Pharmacol. Rep.* **2017**, *69*, 1328–1340. [CrossRef]
220. Wolverton, D.; Blair, M.M. Fracture risk associated with common medications used in treating type 2 diabetes mellitus. *Am. J. Heal. Pharm.* **2017**, *74*, 1143–1151. [CrossRef]
221. Lecka-Czernik, B. Bone Loss in Diabetes: Use of Antidiabetic Thiazolidinediones and Secondary Osteoporosis. *Curr. Osteoporos. Rep.* **2010**, *8*, 178–184. [CrossRef]
222. Yan, W.; Li, X. Impact of diabetes and its treatments on skeletal diseases. *Front. Med.* **2013**, *7*, 81–90. [CrossRef] [PubMed]
223. Guja, C.; Guja, L.; Miulescu, R.D. Effect of type 2 diabetes medications on fracture risk. *Ann. Transl. Med.* **2019**, *7*, 580. [CrossRef] [PubMed]

224. Shirinsky, I.; Shirinsky, V.S. Effects of medication-treated diabetes on incidence and progression of knee osteoarthritis: A longitudinal analysis of the Osteoarthritis Initiative data. *Rheumatol. Int.* **2017**, *37*, 983–991. [CrossRef] [PubMed]
225. Gibbons, C.L.M.H.; Petra, M.; Smith, R.; Athanasou, N.A. Bisphosphonate Treatment of Benign Multifocal and Unifocal Osteolytic Tumours of Bone. *Sarcoma* **2003**, *7*, 35–41. [CrossRef]
226. Balke, M.; Campanacci, L.; Gebert, C.; Picci, P.; Gibbons, M.; Taylor, R.; Hogendoorn, P.; Kroep, J.; Wass, J.; Athanasou, N. Bisphosphonate treatment of aggressive primary, recurrent and metastatic Giant Cell Tumour of Bone. *BMC Cancer* **2010**, *10*, 462. [CrossRef] [PubMed]
227. Gupta, M.; Gupta, N. *Bisphosphonate Related Jaw Osteonecrosis*; StatPearls Publishing: Treasure Island, FL, USA, 2021. Available online: https://www.ncbi.nlm.nih.gov/books/NBK534771/ (accessed on 1 March 2022).
228. Eleutherakis-Papaiakovou, E.; Bamias, A. Antiresorptive treatment-associated ONJ. *Eur. J. Cancer Care* **2017**, *26*, e12787. [CrossRef] [PubMed]
229. Marx, R.E.; Sawatari, Y.; Fortin, M.; Broumand, V. Bisphosphonate-Induced Exposed Bone (Osteonecrosis/Osteopetrosis) of the Jaws: Risk Factors, Recognition, Prevention, and Treatment. *J. Oral Maxillofac. Surg.* **2005**, *63*, 1567–1575. [CrossRef]
230. Siddiqi, A.; Payne, A.G.T.; Zafar, S. Bisphosphonate-induced osteonecrosis of the jaw: A medical enigma? *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endodontol.* **2009**, *108*, e1–e8. [CrossRef]
231. Lo, J.C.; O'Ryan, F.S.; Gordon, N.P.; Yang, J.; Hui, R.L.; Martin, D.; Hutchinson, M.; Lathon, P.V.; Sanchez, G.; Silver, P.; et al. Prevalence of Osteonecrosis of the Jaw in Patients With Oral Bisphosphonate Exposure. *J. Oral Maxillofac. Surg.* **2010**, *68*, 243–253. [CrossRef]
232. Vidal-Real, C.; Sayáns, M.P.; Suarez-Penaranda, J.; Gandara-Rey, J.; Garcia-Garcia, A. Osteonecrosis of the jaws in 194 patients who have undergone intravenous bisphosphonate therapy in Spain. *Medicina Oral Patología Oral y Cirugia Bucal* **2015**, *20*, e267–e272. [CrossRef]
233. Favia, G.; Tempesta, A.; Limongelli, L.; Crincoli, V.; Maiorano, E. Medication-Related Osteonecrosis of the Jaws: Considerations on a New Antiresorptive Therapy (Denosumab) and Treatment Outcome after a 13-Year Experience. *Int. J. Dent.* **2016**, *2016*, 1801676. [CrossRef] [PubMed]
234. Simon, J.A.; Recknor, C.; Moffett, A.H., Jr.; Adachi, J.D.; Franek, E.; Lewiecki, E.M.; McClung, M.R.; Mautalen, C.A.; Ragi-Eis, S.; Nicholson, G.C.; et al. Impact of denosumab on the peripheral skeleton of postmenopausal women with osteoporosis: Bone density, mass, and strength of the radius, and wrist fracture. *Menopause* **2013**, *20*, 130–137. [PubMed]
235. Cosman, F. Anabolic and Antiresorptive Therapy for Osteoporosis: Combination and Sequential Approaches. *Curr. Osteoporos. Rep.* **2014**, *12*, 385–395. [CrossRef] [PubMed]
236. Dadwal, G.; Schulte-Huxel, T.; Kolb, G. Effect of antithrombotic drugs on bone health. *Zeitschrift für Gerontologie und Geriatrie* **2019**, *53*, 457–462. [CrossRef] [PubMed]
237. Signorelli, S.S.; Scuto, S.; Marino, E.; Giusti, M.; Xourafa, A.; Gaudio, A. Anticoagulants and Osteoporosis. *Int. J. Mol. Sci.* **2019**, *20*, 5275. [CrossRef]
238. Gage, B.F.; Birman-Deych, E.; Radford, M.J.; Nilasena, D.S.; Binder, E.F. Risk of osteoporotic fracture in elderly patients taking warfarin: Results from the National Registry of Atrial Fibrillation 2. *Arch. Intern. Med.* **2006**, *166*, 241–246. [CrossRef]
239. Lindner, T.; Cockbain, A.J.; A El Masry, M.; Katonis, P.; Tsiridis, E.; Schizas, C.; Tsiridis, E. The effect of anticoagulant pharmacotherapy on fracture healing. *Expert Opin. Pharmacother.* **2008**, *9*, 1169–1187. [CrossRef]
240. Abdulameer, A.H.; Sulaiman, S.; Kader, M. An Assessment of Osteoporotic Conditions among Users and Non-Users of Warfarin: A Case-Control Study. *J. Clin. Diagn. Res.* **2017**, *11*, OC21–OC24. [CrossRef]
241. Lester, J.; Coleman, R. Bone loss and the aromatase inhibitors. *Br. J. Cancer* **2005**, *93*, S16–S22. [CrossRef]
242. Perez, E.A.; Weilbaecher, K. Aromatase inhibitors and bone loss. *Oncology* **2006**, *20*, 1029–1048.
243. Ribot, C.; Tremollieres, F.; Pouilles, J.M.; Louvet, J.P. Bone mineral density and thyroid hormone therapy. *Clin. Endocrinol.* **1990**, *33*, 143–154. [CrossRef] [PubMed]
244. Uzzan, B.; Campos, J.; Cucherat, M.; Nony, P.; Boissel, J.P.; Perret, G.Y. Effects on bone mass of long term treatment with thyroid hormones: A meta-analysis. *J. Clin. Endocrinol. Metab.* **1996**, *81*, 4278–4289. [CrossRef] [PubMed]
245. Wexler, J.A.; Sharretts, J. Thyroid and bone. *Endocrinol. Metab. Clin. N. Am.* **2007**, *36*, 673–705. [CrossRef]
246. Vestergaard, P. Skeletal Effects of Drugs to Treat Cancer. *Curr. Drug Saf.* **2008**, *3*, 173–177. [CrossRef]
247. Monroy-Cisneros, K.; Esparza-Romero, J.; Valencia, M.E.; Guevara-Torres, A.G.; Méndez-Estrada, R.O.; Anduro-Corona, I.; Astiazarán-García, H. Antineoplastic treatment effect on bone mineral density in Mexican breast cancer patients. *BMC Cancer* **2016**, *16*, 1–7. [CrossRef]
248. Bodensteiner, D.C.; Doolittle, G.C. Adverse Haematological Complications of Anticancer Drugs. *Drug Saf.* **1993**, *8*, 213–224. [CrossRef]
249. Madeddu, C.; Neri, M.; Sanna, E.; Oppi, S.; Macciò, A. Experimental Drugs for Chemotherapy- and Cancer-Related Anemia. *J. Exp. Pharmacol.* **2021**, *13*, 593–611. [CrossRef] [PubMed]
250. Poulopoulos, A.; Papadopoulos, P.; Andreadis, D. Chemotherapy: Oral side effects and dental interventions. A review of the literature. *Stomatol. Dis. Sci.* **2017**, *1*, 35–49. [CrossRef]
251. Naidoo, U.; Goff, D.; Klibanski, A. Hyperprolactinemia and bone mineral density: The potential impact of antipsychotic agents. *Psychoneuroendocrinology* **2003**, *28*, 97–108. [CrossRef]

252. Aronson, J.K. Chapter 6—Antipsychotic Drugs. In *A Worldwide Yearly Survey of New Data in Adverse Drug Reactions and Interactions*; Side Effects of Drugs Annual Book Series, 35; Aronson, J.L., Ed.; Elsevier: Oxford, UK, 2014; pp. 85–131.
253. Shetty, V.; Harrell, L.; Murphy, D.A.; Vitero, S.; Gutierrez, A.; Belin, T.R.; Dye, B.A.; Spolsky, V.W. Dental disease patterns in methamphetamine users: Findings in a large urban sample. *J. Am. Dent. Assoc.* **2015**, *146*, 875–885. [CrossRef]
254. Lee, H.H.; Sudhakara, P.; Desai, S.; Miranda, K.; Martinez, L.R. Understanding the Basis of METH Mouth Using a Rodent Model of Methamphetamine Injection, Sugar Consumption, and Streptococcus mutans Infection. *mBio* **2021**, *12*. [CrossRef]
255. Antoniazzi, R.P.; Zanatta, F.B.; Rösing, C.K.; Feldens, C.A. Association Among Periodontitis and the Use of Crack Cocaine and Other Illicit Drugs. *J. Periodontol.* **2016**, *87*, 1396–1405. [CrossRef] [PubMed]
256. Antoniazzi, R.P.; Sari, A.R.; Casarin, M.; De Moraes, C.M.B.; Feldens, C.A. Association between crack cocaine use and reduced salivary flow. *Braz. Oral Res.* **2017**, *31*, e42. [CrossRef] [PubMed]
257. Blanksma, C.J.; Brand, H.S. Cocaine abuse: Orofacial manifestations and implications for dental treatment. *Int. Dent. J.* **2005**, *55*, 365–369. [CrossRef] [PubMed]
258. Brand, H.S.; Gonggrijp, S.; Blanksma, C.J. Cocaine and oral health. *Br. Dent. J.* **2008**, *204*, 365–369. [CrossRef] [PubMed]
259. Cury, P.R.; Oliveira, M.G.A.; Santos, J. Periodontal status in crack and cocaine addicted men: A cross-sectional study. *Environ. Sci. Pollut. Res.* **2016**, *24*, 3423–3429. [CrossRef]
260. Keboa, M.T.; Enriquez, N.; Martel, M.; Nicolau, B.; Macdonald, M.E. Oral Health Implications of Cannabis Smoking: A Rapid Evidence Review. *J. Can. Dent. Assoc.* **2020**, *86*, k2.
261. Thomas, M.; Denny, C. Medication-related tooth discoloration: A review. *Dent. Updat.* **2014**, *41*, 440–447. [CrossRef]
262. Ashukem, M.T.; Levy, J.C.; Formaini, N. Minocycline induced black bone disease: An incidental finding during total shoulder arthroplasty. *Curr. Orthop. Pr.* **2016**, *27*, 698–701. [CrossRef]
263. Sánchez, R.A.; Rogers, S.R.; Sheridan, J.P. Tetracycline and other tetracycline—Derivative staining of the teeth and oral cavity. *Int. J. Dermatol.* **2004**, *43*, 709–715. [CrossRef]
264. Alkhatib, M.N.; Holt, R.D.; Bedi, R. Smoking and tooth discolouration: Findings from a national cross-sectional study. *BMC Public Heal.* **2005**, *5*. [CrossRef] [PubMed]
265. Watts, A.; Addy, M. Tooth discolouration and staining: A review of the literature. *Br. Dent. J.* **2001**, *190*, 309–316. [CrossRef] [PubMed]
266. Rhomberg, W.; Schuster, A. Premature Calcifications of Costal Cartilages: A New Perspective. *Radiol. Res. Pr.* **2014**, *2014*, 1–5. [CrossRef] [PubMed]
267. Felson, D.T.; Ms, G.R.; Chen, X.; LaValley, M.; Jafarzadeh, S.R.; Lewis, C.E.; Torner, J.; Nevitt, M.C.; Misra, D.; MOST Study Investigators. The Association of Diuretics and Proton Pump Inhibitors With Chondrocalcinosis. *ACR Open Rheumatol.* **2021**, *3*, 390–394. [CrossRef]
268. Gotway, M.B.; Marder, S.R.; Hanks, D.K.; Leung, J.W.T.; Dawn, S.K.; Gean, A.D.; Reddy, G.P.; Araoz, P.A.; Webb, W.R. Thoracic Complications of Illicit Drug Use: An Organ System Approach. *Radiographics* **2002**, *22*, S119–S135. [CrossRef]
269. Stone, S.; A Malanga, G.; Capella, T. Corticosteroids: Review of the History, the Effectiveness, and Adverse Effects in the Treatment of Joint Pain. *Pain Physician* **2021**, *24*, S233–S246.
270. Huang, Y.-J.; Kuo, C.-F. Can drugs trigger CPPD acute attacks? *Jt. Bone Spine* **2018**, *86*, 131–134. [CrossRef]
271. Matlaga, B.R.; Shah, O.D.; Assimos, D.G. Drug-induced urinary calculi. *Rev. Urol.* **2003**, *5*, 227–231.
272. Daudon, M.; Jungers, P. Drug-induced renal calculi: Epidemiology, prevention and management. *Drugs* **2004**, *64*, 245–275. [CrossRef]
273. Franceschetti, L.; Di Candia, D.; Giordano, G.; Carabelli, I.; Vignali, G.; Cattaneo, C. Drugs in bone: Detectability of substances of toxicological interest in different states of preservation. *J. Forensic Sci.* **2021**, *66*, 677–686. [CrossRef]
274. Fernandez-Lopez, L.; Mancini, R.; Pellegrini, M.; Rotolo, M.C.; Luna, A.; Falcon, M. Postmortem analysis of quetiapine and pregabalin in human bone. *Leg. Med.* **2020**, *46*, 101717. [CrossRef] [PubMed]
275. McIntyre, I.M.; King, C.V.; Boratto, M.; Drummer, O.H. Post-Mortem Drug Analyses in Bone and Bone Marrow. *Ther. Drug Monit.* **2000**, *22*, 79–83. [CrossRef] [PubMed]
276. Fernandez-Lopez, L.; Pellegrini, M.; Rotolo, M.C.; Maldonado, A.L.; Falcon, M.; Mancini, R. Development and validation of a method for analysing of duloxetine, venlafaxine and amitriptyline in human bone. *Forensic Sci. Int.* **2019**, *299*, 154–160. [CrossRef] [PubMed]
277. Fernandez-Lopez, L.; Pellegrini, M.; Rotolo, M.C.; Luna, A.; Falcon, M.; Mancini, R. Development and Validation of a Method for the Analysis of Bisoprolol and Atenolol in Human Bone. *Molecules* **2019**, *24*, 2400. [CrossRef] [PubMed]
278. Fernandez-Lopez, L.; Luna-Maldonado, A.; Falcon, M.; Mastrobattista, L.; Navarro-Zaragoza, J.; Mancini, R. Development and validation of a gas chromatography–mass spectrometry method for opiates and cocaine in human bone. *J. Pharm. Biomed. Anal.* **2019**, *164*, 636–641. [CrossRef]
279. Vardakou, I.; Athanaselis, S.; Pistos, C.; Papadodima, S.; Spiliopoulou, C.; Moraitis, K. The clavicle bone as an alternative matrix in forensic toxicological analysis. *J. Forensic Leg. Med.* **2014**, *22*, 7–9. [CrossRef]
280. Raikos, N.; Tsoukali, H.; Njau, S. Determination of opiates in postmortem bone and bone marrow. *Forensic Sci. Int.* **2001**, *123*, 140–141. [CrossRef]
281. Fitzpatrick, L.A. Secondary causes of osteoporosis. *Mayo Clin. Proc.* **2002**, *77*, 453–468. [CrossRef]

biology

MDPI

Article

Age-at-Death Estimation of Fetuses and Infants in Forensic Anthropology: A New "Coupling" Method to Detect Biases Due to Altered Growth Trajectories

Mélissa Niel [1,*], Kathia Chaumoître [1,2] and Pascal Adalian [1]

[1] Aix Marseille Univ, CNRS, EFS, ADES, 13007 Marseille, France; kathia.chaumoitre@univ-amu.fr (K.C.); pascal.adalian@univ-amu.fr (P.A.)
[2] Service de Radiologie et Imagerie Médicale, Hôpital Nord, CHU Marseille, Assistance Publique des Hôpitaux de Marseille, 13015 Marseille, France
* Correspondence: melissaniel@hotmail.fr; Tel.: +33-4-9169-8890

Simple Summary: In forensic anthropology, estimating the age-at-death of young juvenile skeletons is crucial as a direct determinant of legal issues in many countries. Most methods published for this purpose are based on either maturation or growth processes (two essential components of development) and focus on "normal" (i.e., nonpathological) growth. However, when the osseous remains available for study are from an individual that experienced an altered growth process, age estimation may be biased, and accounting for this would be helpful for potentially avoiding inaccuracies in estimation. In this research, we developed a method based on the combined evaluation of both maturation and growth. Maturation is evaluated by the conformation of the *pars basilaris*, a bone at the skull base that provides an indirect estimate of brain maturation, while growth is assessed using femoral biometry. The method was tested on two medical validation samples of normal and pathological individuals. The results show that it was possible to identify "uncoupling" between maturation and growth in 22.8% of the pathological individuals. Highlighting potential uncoupling is therefore an essential step in assessing the confidence of an age estimate, and its presence should lead experts to be cautious in their conclusions in court.

Abstract: The coupling between maturation and growth in the age estimation of young individuals with altered growth processes was analyzed in this study, whereby the age was determined using a geometric morphometrics method. A medical sample comprising 223 fetuses and infants was used to establish the method. The *pars basilaris* shapes, quantified by elliptic Fourier analysis, were grouped into consensus stages to characterize the maturation process along increasing age groups. Each *pars basilaris* maturation stage was "coupled" to biometry by defining an associated femur length range. The method was tested on a validation sample of 42 normal individuals and a pathological sample of 114 individuals whose pathologies were medically assessed. Couplings were present in 90.48% of the normal sample and 77.19% of the pathological sample. The method was able to detect "uncoupling" (i.e., possibly altered growth) in more than 22.8% of samples, even if there was no visible traces of pathology on bones in most cases. In conclusion, experts should be warned that living conditions may cause alterations in the development of young individuals in terms of uncoupling, and that the age-at-death estimation based on long bone biometry could be biased. In a forensic context, when age has been estimated in cases where uncoupling is present, experts should be careful to take potential inaccuracies into account when forming their conclusions.

Keywords: forensic anthropology; age estimation; femur length; *pars basilaris* shape; inverse Fourier transform; geometric morphometrics

Citation: Niel, M.; Chaumoître, K.; Adalian, P. Age-at-Death Estimation of Fetuses and Infants in Forensic Anthropology: A New "Coupling" Method to Detect Biases Due to Altered Growth Trajectories. *Biology* **2022**, *11*, 200. https://doi.org/10.3390/biology11020200

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 23 December 2021
Accepted: 25 January 2022
Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Estimating an individual's age-at-death from skeletal remains is one of the major issues in biological and forensic anthropology when assessing a biological profile. In the case of young individual skeletons, age-at-death is crucial to any analysis of biological remains. In forensic anthropology, a fetus's legal personality is dependent on fetal age estimation, with the resulting social, ethical, and economic consequences [1], and the assessment of fetal viability and legislation on abortion and infanticide are also directly dependent on fetal and infant age estimation—hence contributing to the need and importance of developing reliable and accurate methods.

Several fetal and infant age-at-death estimation methods have been established. Most of these are osteometric, radiographic, or ultrasound methods [2–20]. They can be development based, which aim to estimate physiological age based on maturation processes (e.g., skeletal morphology, appearance and maturation of secondary ossification centers, maturation of dental germs), or biometric based, which rely on growth processes (e.g., crown–rump length, cranial and abdominal perimeters, and the maximum length of long bones).

However, the question of living conditions and, therefore, the context in which the development of a young juvenile took place can remain unanswered. Most methods assume that these conditions are "favorable" or "normal", though these can obviously be disturbed by any pathological conditions experienced by the mother or child. In other words, the ontogenetic trajectory—the child's developmental trajectory—is likely to be altered.

It is generally accepted that brain maturation is the best criterion to establish physiological age during early development, regardless of the environmental or socioeconomic conditions, even in cases of fetal or maternal pathologies [21–23]. The brain unfortunately undergoes rapid autolysis after death (within approximately 48 h) and can no longer be studied, but it has an influence on skull base osseous structures [24–31]. Therefore, these structures can be considered to be indirect and taphonomically resistant testaments of brain maturation.

To establish a biometric age, it is accepted that femoral length is the most reliable and accurate estimation indicator [3,7–9,32]. Nevertheless, growth-based age estimation may be biased in cases of growth delay or growth advancement caused by pathological conditions. These conditions are difficult to detect because most pathologies leave little or no trace on fetal and infant bones. Sherwood et al. [32] demonstrated that diseases causing abnormally short femurs (such as trisomy 21 or Turner syndrome) or abnormally long femurs (such as *spina bifida*) can lead to inaccuracies of up to almost four weeks in fetuses when estimating age at death.

Therefore, when only using femoral length without considering possible alterations in developmental conditions, one cannot know whether the age at death will be underestimated, correct, or overestimated with respect to the chronological age (real age).

Our biological hypothesis is that the physiological age (maturation) is more reliable and stable than the biometric age (growth), and that these two "different kinds of ages" are coupled for nonpathological individuals. Accepting this hypothesis, it can be argued that living conditions, whether they are simply "changing" or truly "unfavorable" to development, influence biometric growth more than maturation.

This "coupling" or agreement between maturation and growth processes could be used to assess and control fetal and infant age-at-death estimation, targeting individuals with growth variation due to possible pathological conditions. As a consequence, the demonstration of the "uncoupling" of these two processes would be an indication, or even serve as an alert, that the accuracy of the age-at-death estimation of a young juvenile skeleton must be considered with great caution.

As a direct indicator of skull base maturation and therefore an indirect indicator of brain (and thus general) maturation, we chose to use the *pars basilaris* of the future occipital bone [33–36]. We quantified its degree of maturation with geometric morphometric analyses

from its outline shape. The estimation of biometric age (growth) was based on the maximum diaphyseal length of the femur.

These two bones are both dense and compact [11,37,38], and they are generally found to be in good preservation states considering forensic and archaeological contexts [11,37].

Using computerized tomography (CT) scan imaging of fetuses and infants with nonpathological conditions, the aim of our study was to develop a method based on the expected coupling between maturation and growth to detect possible growth variation.

Once established on a medical imaging sample (learning sample) of nonpathological individuals, the method was applied to a separate validation medical sample of nonpathological individuals and another sample of individuals whose pathologies were fully documented.

If an individual presents the "normal" (i.e., nonpathological) coupling variability established by the learning sample, the hypothesis of an alteration of his ontogenetic trajectory can be proposed. It is then necessary to discuss the potential reason for this alteration (growth delay or advancement in connection or otherwise with an identified pathology). Regardless, this study shows that estimated age must be considered with caution.

2. Materials and Methods

2.1. Sample

An anonymized database composed of 1136 individuals aged between 11 weeks in utero and 20 years old was compiled within UMR 7268 ADES (AMU-CNRS-EFS). From this, a medical imaging sample of 379 individuals aged 16 weeks in utero to approximately one and a half years (17.7 months) was derived.

2.1.1. Normal and Pathological Development

The studied population was divided into three samples. A learning sample (A) comprising 223 fetuses and infants with nonpathological conditions (77 girls, 115 boys, and 31 of unknown sex) ranging from 16 fetal weeks to 77 postnatal weeks (mean age: 33.28 fetal weeks; Figure 1) was used to establish the method. A second sample (B) comprising 42 fetuses and infants ranging between 18 fetal weeks and 61 postnatal weeks (mean age: 34.69 fetal weeks; Figure 1) was used as a separate validation sample. Given that the available age classes were not homogeneous for normal individuals, random selection by age classes was conducted to ensure a good representation of age; the selection comprised approximately 85% for the learning sample and 15% for the validation sample.

For our analyses, the ages of fetuses (based on accurate reports of the mother's last normal menstrual period and ultrasound data obtained at 10 weeks of gestation, which is an obligatory examination under French law) and infants were expressed in weeks: from 16 to 38 weeks for fetuses and from 39 to 115 weeks for postnatal individuals. This means that a "45-week-old" individual is actually an individual aged 45 weeks minus 38 weeks (average length of pregnancy), which corresponds to 7 postnatal weeks.

Nonpathological conditions were essential for sample A and B individuals. The conditions considered for mothers were the absence of congenital disease, diabetes, or arterial hypertension. The nonpathological conditions of fetuses (such as the absence of external or visceral malformation, the absence of bone anomaly on a CT scan, the absence of cerebral anomaly on MRI, and normal karyotypes) were established based on multidisciplinary ante mortem and post mortem examinations conducted by medical experts of the prenatal diagnosis center. Concerning infants, CT scans allowed us to verify developmental normality. Examinations were performed in cases of road accidents, sudden or unexpected infant death syndrome, and forensic investigations.

Fetuses and infants with identified pathological conditions were included in a third sample (C) comprising 114 fetuses and infants (61 girls, 47 boys, and 6 of unknown sex) ranging from 16 fetal weeks to 47 postnatal weeks (mean age: 27.24 fetal weeks) (Figure 1C).

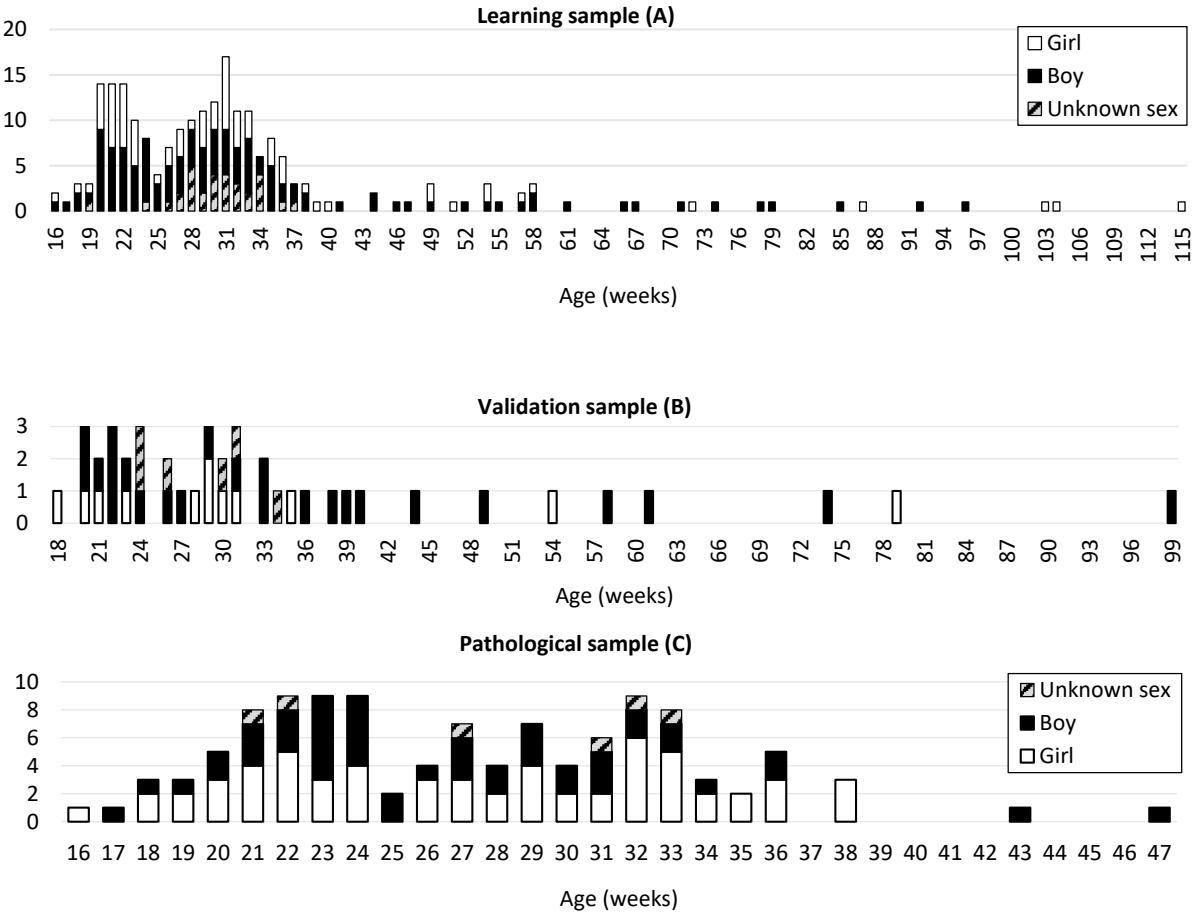


Figure 1. Age (in weeks) and sex distribution of the learning sample (**A**) comprising 223 individuals and the validation sample (**B**) comprising 42 individuals. Age (in weeks) and sex distribution of the pathological sample (**C**) comprising 114 individuals.

2.1.2. Pathologies Groups

Depending on the pathological conditions, the following subgroups were established:

- Constitutional bone diseases or CBD (Ellis–van Creveld syndrome, thanatophoric dysplasia, achondroplasia, Jeune syndrome, facial femoral syndrome, VACTERL association, and harlequin ichthyosis = 14%);
- Growth disorders or conditions justifying differentiated growth or GD (intrauterine growth retardation, macrosomia/diabetes, and twin pregnancy = 39%);
- Localized anomalies or LA (skull, polymalformative syndrome, limbs, and spine = 23%);
- Cerebral anomalies or CA (21%);
- Chromosomal anomalies or CHRA (trisomy 21 and trisomy 18 = 3%).

The same individual could be classified in several types of pathologies, such as a localized anomaly and a cerebral anomaly.

2.2. Data Acquisition

The ante mortem and post mortem CT scans of sample A, B, and C individuals were collected from the Picture Archiving and Communication System (PACS) in the hospital of Marseilles (Assistance Publique—Hôpitaux de Marseille, France). Individuals were scanned using a helical CT scanner (Somatom Sensation Cardiac 64; Siemens, Erlangen, Germany). The scanning parameters were as follows: voltage of 100–140 kVp, amperage of 50–180 mAs, 512 × 512 pixels, resolution of 0.25–4.87 pixels per mm, voxel size of approximatively 0.5 × 0.5 × 0.6 or 1 mm^3, and a slice thickness of 0.6–1 mm. These high-resolution native slices recorded in the Digital Imaging and Communications in Medicine

(DICOM) format were anonymized before being used in the study, in accordance with the standards of the French National Consultative Ethics Committee for health and life sciences (CCNE) and the Helsinki Declaration of 1975 concerning the privacy and confidentiality of personal data.

2.3. Bone Reconstruction

Before reconstructing the femur and *pars basilaris* in three dimensions (3D), region of interest (ROI) segmentation on the DICOM slices was performed with the ImageJ®v1.51 software (National Institutes of Health, Bethesda, MD, USA) to separate the bone from adjacent tissues. The threshold value was obtained by calculating a threshold mean value (TMV) [38], which is an average of the half-maximum height (HMH) values [39]. The TMV was used in Avizo Standard Edition (v.7.0.0®, Visualization Sciences Group, SAS, Berlin, Germany) to reconstruct the 3D bone surfaces.

Since there are no significant differences between the right and left femur in young juveniles [3,9,10,40–42] and convention suggests that the left femur is preferred, we only measured the right femur when the left was not available.

2.4. Maturation Criterion: Elliptic Fourier Analysis of the Pars Basilaris

The complete protocol was described by Niel et al. [43] and was used in this study.

2.4.1. Outline Process

Briefly, we defined a homologous reference plane for all the *pars basilaris* in the inferior (external) view. This was defined thanks to two type II and one type III landmarks [44]. Type II landmarks are the most posterior point of the left and right horns, and a type III landmark is the central point of the anterior surface. All landmarks were digitized on 3D reconstructed surfaces using Avizo Standard Edition® software.

This step allowed us to project all reconstructions in the same 2D plane and with the same orientation. Then, outline shapes were quantified according to 150 equally linearly spaced points digitized along the *pars basilaris* contour with the tpsDig2 v.2.17® digitization program [45]. Finally, the contour data of the *pars basilaris* were normalized using generalized Procrustes analysis (GPA) [46–49] based on four type II and III homologous landmarks [44] called control points [46].

2.4.2. Measurement Error

Repeatability (intra-observer error) and reproducibility tests (inter-observer error) were realized to validate the protocol on 30 randomly selected individuals in sample A. Repeatability was tested by the same observer repeating the protocol twice several weeks apart; for reproducibility, a second observer applied the protocol once.

2.4.3. Harmonics Number

With EFA, one may wonder what the appropriate number of harmonics is, since this number determines the accuracy of the contour reconstruction. The following two paragraphs of text is the explanation as reproduced from Niel et al. pp. 37–38 [43]:

According to the Nyquist theorem [50], the harmonic number must be less than half the number of sampled outline points. Consequently, on the 150 points sampled for EFA, only the first 74 harmonics were retained for analysis. Given that we cannot retain all the Fourier coefficients for our analysis" (74 harmonics $\times$ 4 coefficients = 296 coefficients), because the measurement error is expected to increase with harmonic ranks, the percentage of error on harmonic coefficients was calculated using a Procrustes analysis of variance (ANOVA) on the three sessions [51]. This procedure calculated the mean sums of squares for the four coefficients of each harmonic to observe the evolution of error according to the rank of the harmonics (in percentage). Only the first harmonics, showing an acceptable digitization error rate, were retained for further analyses. An error rate under 35% is considered to be reasonable in an outline analysis using EFA [51].

The assessment of the total percentage of measurement error was then performed using a Procrustes ANOVA [51–55] adapted to elliptic Fourier coefficients [51]. The Fourier coefficients of the coupled series are used in the Procrustes ANOVA with the number of harmonics previously defined. The intra- and interindividual variances were directly calculated from the means of the sums of squares and crossed products corresponding to individuals and residual sources of variation [51]. These residuals, representing the variability between the two sessions, correspond to the measurement error [55].

2.5. Coupling between Maturation and Growth Process

2.5.1. Maturation Criterion: Shape Stages

Maturation stages were established on the *pars basilaris* shapes of the nonpathological learning sample A to visualize the *pars basilaris* morphological changes through time. With this sample, consensus shapes from 4 to 26 weeks with overlap every 2 to 13 weeks were created, which enabled us to have intermediate shapes and, thus, a continuous vision of maturation from 16 weeks in utero to 77 postnatal weeks. Thus, 19 stages of consensus shapes, defined by the mean of 5–52 shapes depending on stages, were obtained (Table 1). Then, to visualize and compare the morphology of each consensus shape, the *pars basilaris* outlines were reconstructed from Fourier coefficients with the inverse Fourier transform function [56–58].

Table 1. Sample A: age group (in weeks), number of weeks, and number of individuals according to the 19 *pars basilaris* maturation stages as well as femoral growth in percentiles (minimal values of 0–10, 10, 50, and 90 and maximal value of 100 + 10, in millimeters).

Stage	Age Group (Weeks)	Number of Weeks	Number of Individuals	Percentiles				
				0–10	10	50	90	100 + 10
1	16–19	4	9	22.11	25.57	30.09	35.33	40.08
2	18–21	4	34	23.41	31.58	37.93	41.12	46.34
3	20–23	4	52	27.71	34.73	39.42	44.36	51.21
4	22–25	4	36	32.25	36.88	43.19	47.18	54.30
5	24–27	4	28	33.41	43.17	49.34	54.97	58.60
6	26–29	4	37	38.74	48.04	52.20	56.72	63.89
7	28–31	4	50	43.60	51.15	56.70	60.42	65.57
8	30–33	4	51	46.62	56.22	59.28	62.54	72.43
9	32–35	4	36	51.58	58.09	62.53	66.44	81.15
10	34–37	4	23	61.72	63.59	65.56	71.68	75.91
11	36–40	5	14	61.55	65.66	71.14	76.48	79.70
12	38–44	7	8	71.60	73.41	77.71	88.29	102.87
13	41–51	11	9	79.18	82.26	90.68	94.95	96.20
14	44–57	14	15	75.98	85.46	94.79	104.14	105.43
15	51–66	16	13	87.78	94.42	103.51	111.75	122.13
16	58–78	21	10	97.49	107.55	112.40	117.30	118.52
17	67–92	26	9	104.94	111.99	117.91	135.63	164.87
18	79–104	26	7	121.72	123.99	132.02	151.45	155.06
19	92–115	24	5	131.92	132.13	149.79	152.05	154.46

2.5.2. Growth Criterion: Femoral Lengths

Femoral diaphysis lengths were measured (in millimeters) on Avizo Standard Edition®. Percentiles were calculated from sample A according to each maturation stage of the *pars basilaris* and used as growth criteria (Table 1). To include a greater range, a margin of ten percentiles was added at each extreme, calculated as the difference between 0 and 10 percentiles and between 100 and 90 percentiles, thus providing 0–10 percentiles and 100 + 10 percentiles, respectively (Table 1).

2.6. Statistical and Morphometric Analyses

2.6.1. Bilateral Femoral Asymmetry and Sex Effect on the Variables

Between-sex comparisons of the *pars basilaris* shapes were explored using nonparametric multivariate ANOVA (MANOVA) [59], and the between-sex comparison of the femoral lengths was performed using Kruskal–Wallis rank sum testing. The bilateral femoral asymmetry was explored using a *t*-test.

2.6.2. Application of the Coupling Method in Samples B and C

Each *pars basilaris* of samples B and C was tested, one at a time, by comparison with the 19 stages representing the maturation consensus shapes. Once the outlines were quantified with EFA after the GPA procedure, assigning a maturation stage to the tested *pars basilaris* was realized by calculating the Euclidian distance (or Procrustes distance) between the centroids of the 19 consensus stages and the tested (compared) shape [60,61]. The minimal distance between the centroid of the tested *pars basilaris* and one of the 19 consensus shapes allowed for the assignation of a stage to the *pars basilaris*.

For growth, the measurement of the tested individual femoral length was compared to the range expected for the defined maturation stage (Table 1). If this measurement was found to be within the expected range, we considered that growth corresponded to the maturation stage values and there was "coupling". Then, it could be concluded that growth was "normal" (i.e., nonpathological). On the contrary, if growth did not correspond to the maturation stage values, then "uncoupling" had occurred.

Analyses were performed using *RStudio* (developed for R software—Version 1.1.383—® 2009–2017 RStudio, Inc., Boston, United States) and the software packages *Momocs* [62], *Morpho* [63], *geomorph* [64], *car* [65], *gap* [66], *efourier*, and *iefourier* functions [56].

3. Results

3.1. Quantification of Pars Basilaris Shapes

3.1.1. Number of Harmonics

The percentage of measurement error was inferior to the threshold defined at 10% for the first 14 harmonics, corresponding to 56 Fourier coefficients per individual. This allowed us to faithfully reconstruct the outline of the *pars basilaris* (Figure 2).

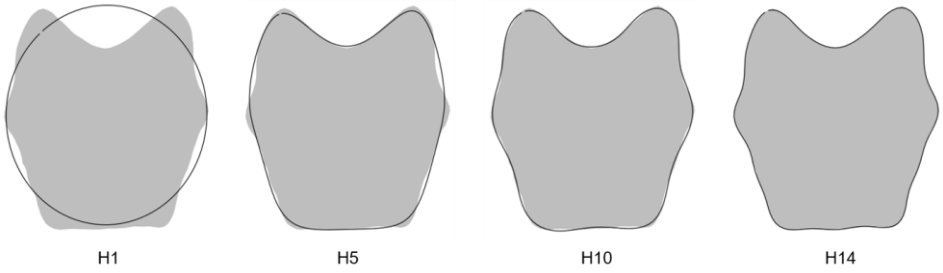


Figure 2. Outline reconstructions of the *pars basilaris* (dark outline): harmonics 1, 5, 10, and 14. Gray shapes represent the reconstruction of the *pars basilaris* with the maximum number of harmonics (74 in this study).

3.1.2. Measurement Error

The percentage of measurement error for the outline protocol is 1.13% for repeatability and 1.96% for reproducibility for the selected first 14 harmonics. This protocol is reliable and reproducible.

3.2. Between-Sex Differences and Femoral Length

The nonparametric MANOVA showed that there were no significant shape differences between sex groups ($F = 1.503$, d$f = 2$, $p = 0.199$) and the femoral lengths were not sig-

nificantly different between sex groups (p = 0.706). Additionally, there was no bilateral asymmetry (p = 0.239) between the right and left femoral diaphysis.

3.3. Coupling between Maturation and Growth

The maturation and growth criteria are summarized in Figure 3. Each maturation stage corresponds to a range of femur lengths defined by the lower bound (0–10 percentiles) and the upper bound (100 + 10 percentiles), corresponding to the extremes.

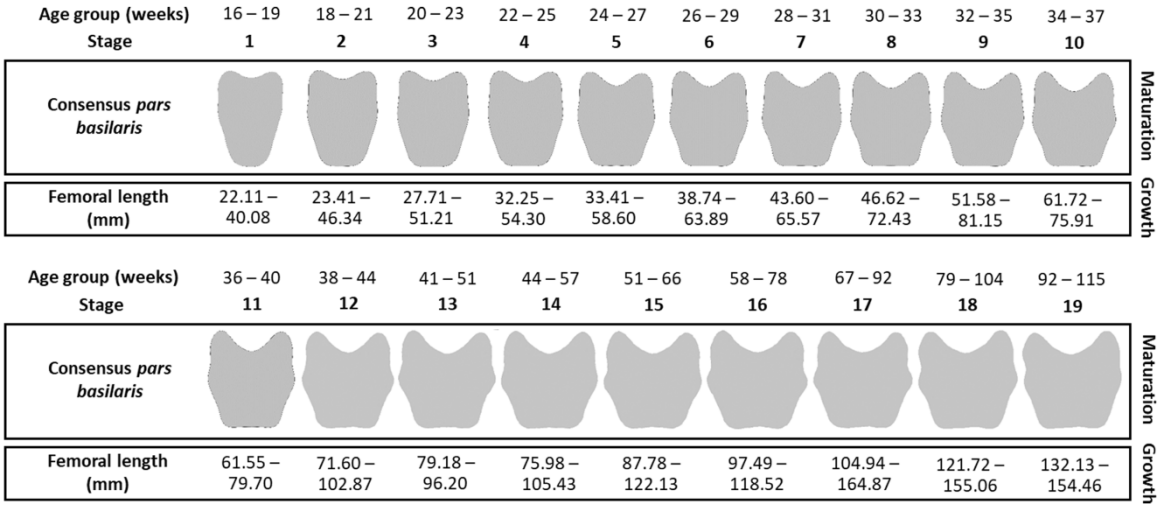


Figure 3. Maturation and growth criteria by stage and age group (in weeks).

Method Application

The method was applied to the validation sample B and the pathological sample C to verify whether growth and maturation were coupled or, in other words, whether the individual's growth corresponded to the values expected by their maturation stage (Figure 3).

In sample B, we observed coupling in 90.48% of samples. The four cases where uncoupling was detected correspond to two growth delays (−4.82 and −1.70 mm) and two growth advancements (+1.26 and +13.13 mm). These values were calculated by subtracting the femoral length of the tested individual (X^T) at the upper (I^s) or lower (I^i) values of the expected interval for the maturation stage, depending on whether individual measurement was inferior or superior to the interval.

For a measurement inferior to the interval:

$$X^T - I^i = -x \text{ or growth delay,}$$

For a measurement superior to the interval:

$$X^T - I^s = +x \text{ or growth advancement.}$$

In sample C, 26 individuals (22.81% of the sample) showed uncoupling. Most of them were girls (61.5%). Uncoupling in these cases corresponded to 14 cases of growth delay (from −23.02 to −1 mm) and 12 cases of growth advancement (from +0.43 to +6.61 mm).

Regarding the subgroups of pathological conditions for uncoupling, LA was the most represented (29%), followed by CBD (26%) and GD (26%) in equal parts; CA was the least represented (19%). More precisely, individuals in the LA subgroup who were most likely to have uncoupling were those presenting a cranial anomaly (45%), followed by polymalformative syndromes (33%) and limb anomalies (11%). In the GD subgroup, IUGR was the most common pathology (50%), followed by macrosomia/diabetes (37%). Then, among individuals with CBD, uncoupling was more frequently observed for the

thanatophoric dysplasia cases (25%) and in relatively equal parts for the other diseases. Finally, CA was the least frequent in uncoupled individuals (19%) (Figure 4).

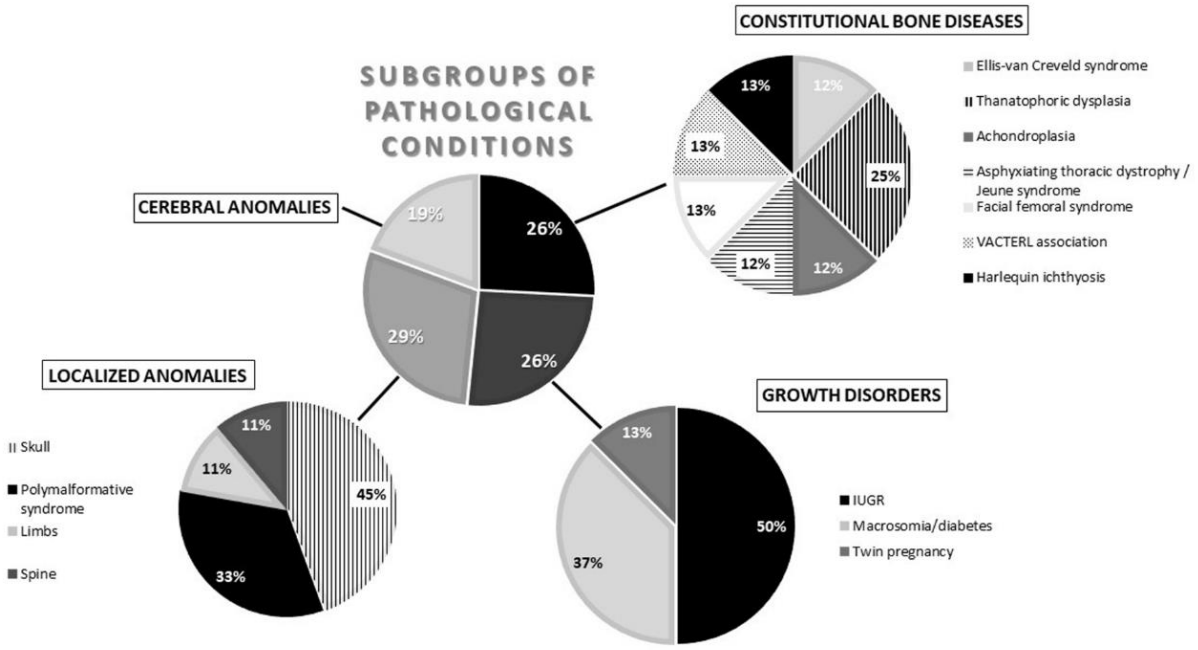


Figure 4. Chart summarizing the subgroups and the detailed pathological conditions for individuals in the medical imaging sample with uncoupling. IUGR = intrauterine growth retardation.

4. Discussion

4.1. The Fetus and Infant Sample

In France, since the advent of prenatal diagnosis centers (Decree 97–578 of 28 May 1997, consolidated on 11 May 2018, France), fetuses have been systematically examined in cases of medically interrupted pregnancy or spontaneous death (miscarriages and in utero deaths). A panel of experts' analyses medical records follows ante mortem (CT scan in utero) and post mortem (complete visceral examination, histological study, fetal karyotype, placenta examination, description of external and visceral abnormalities, and front and profile radiography) examinations. After respecting a strict anonymization protocol, we could access these examinations records and be informed about malformations (bone or visceral), chromosomal abnormalities, or even the precise determination of the cause of death.

For sudden and unexpected infant death and forensic cases, CT scans and autopsies are performed only with the written consent of the parents. Not all parents agreed, which is why there were few available exams. Moreover, sudden and unexpected infant death generally occurs before the age of one year according to the High Authority for Health, which stated in its 2007 report that 80% of sudden infant deaths occur before the age of 6 months, with a peak at 2–3 months. This is consistent with the age distribution of our study sample.

For children aged more than 1 year, we could access some rare autopsy reports and some ante mortem CT scans, which are mostly performed for infants who have fallen or have been in a car accident. Cases are rather rare, and when they exist, the whole body is rarely examined to avoid unnecessary radiation. For our analyses, however, we required images that at least included the portion of the body from the skull base to the proximal end of the tibia. All of these elements made it difficult for us to obtain a large sample of fetuses and infants and almost impossible to have homogeneous age groups.

The second difficulty in studying young individuals concerns the CT scan quality. We first sorted the CT scans according to their image quality as excellent, average, or poor.

This sorting forced us to rule out many examinations that were not exploitable for our study (due to flowing bone surfaces, incomplete bone, and irregular contours). It should be recalled that fetal X-ray exposure increases the risk of malformation (teratogenic effects) and long-term cancer induction (carcinogenic effects) [67], so the dose of radiation should be as low as possible.

In the case of a postmortem CT scan, the dose of radiation may be higher because the same ethical concerns are no longer relevant. These obtained slices were generally of high or excellent quality and therefore represent the largest part of our studied material.

4.2. Quantification of Shape

Geometric morphometric methods and EFA have already been used to quantify the *pars basilaris* shape changes and intrastage variability during the second and third trimesters of fetal life [43]. EFA is suitable for considering the curved morphology and small thickness of this bone, since it is difficult to digitize homologous landmarks on the surface. This difficulty, combined with the fact that the only definable landmarks are not linked to the overall object geometry, oriented us toward a mathematical description of the outline to analyze the global shape of the *pars basilaris.*

As explained by Niel et al., 2019 (pp. 40–41) [43], outline analysis (and, more specifically, Fourier descriptors) provide complex and detailed information regarding the shape. Additionally, this method has been frequently used for discriminating biological forms quantifying morphological differences [46,51,57,58,68–75], as the use of ellipses means that the shape description in EFA is global and therefore helpful for describing bones with curved edges [70,76]. This indicates that it is perfectly suited for characterizing the morphology of the *pars basilaris*.

In the development of the method, a few available landmarks were used to define the reference plane and normalize the Fourier descriptors. The normalization of the control point using GPA [46] prevents the homology problems encountered in specimen alignment on the major axis of the first ellipse, which is conventionally used for the normalization of Fourier descriptors [77]. This method was not adapted to *pars basilaris* because the ratio between the length and width changes as the child develops [11,18,37,78,79]. It has also been shown that among the various normalization methods, the one using the control point with GPA is the most appropriate to use for bones with a few homologous landmarks and circular contours [70,76], such as the *pars basilaris*.

4.3. Interest in the Pars Basilaris

Because of its early formation, between the 10 and 14 gestational weeks [11,78–85], and its robustness, the *pars basilaris* is one of the elements of the future adult occipital bone most used to establish age-at-death estimation methods for fetuses and infants. Methods using this bone generally use conventional morphometry and/or bone size ratio [11,15,18,37,78], but they do not consider the shape, which might be valuable in improving age estimation.

Thanks to geometric morphometric methods based on Cartesian landmark coordinates, some researchers have been interested in shape to document the skull base changes through development, though with no intention of age estimation. Shape is defined as the geometric properties of an object that are invariant to scale, rotation, and translation, whereas the form of an object includes both its shape and size [60,86] (Needham equation: form = shape + size) [87].

Transposed onto our biological or forensic anthropology context concerning bones, the shape corresponds to bone maturation and the size corresponds to growth. The advantage of geometric morphometric methods is their ability to precisely quantify and visualize morphological variation through powerful statistical tools [60,86]. Based on these methods, previous studies have described the fetal cranial base development as a whole [30,36,78,88,89], but the *pars basilaris* morphology has rarely been separately analyzed.

Moreover, most morphometric methods focus on a single anatomical area to estimate age. We believe that the multiplication of age estimators, in addition to increasing re-

liability and accuracy [90], would minimize estimation errors [32,78,91], an idea that is consistent with some previous studies [79,92,93]. For example, according to Tocheri and Molto [91], linear measurements of the *pars basilaris* make it possible to refine the estimated age according to the degree of dental eruption and the maximum length of the femoral diaphysis.

Other studies have shown that femoral length coupled with histological study and the combination of several fetal measurements (biparietal diameter, head circumference, abdominal perimeter, and femur and radius length) improve the accuracy of fetal age estimation [92,93]. Additionally, the *pars basilaris* maximum length is significantly correlated with age, crown–rump length, and humerus length [94]. These studies demonstrate that it is possible to refine age estimation through the use of conventional morphometry together with a combination of several parameters.

4.4. Morphology of the Pars Basilaris

In the literature, several authors have used traditional morphometry to demonstrate that the *pars basilaris* dimensions evolve during fetal and infant development [18,23,78,79], and the bone characteristics intensify with age [23]. The morphological characteristics of the *pars basilaris* are used not only in anatomy but also in biological anthropology, as they can give an idea about the fetal and infant age [11,15,18,37,78,79].

Using geometric morphometric methods, shape analysis confirms the increase in morphological changes from 18 to 41 gestational weeks [43]. The conclusions of our own study allow researchers to precisely quantify and visualize shape changes of the whole *pars basilaris* during prenatal development and after birth for the first time.

By studying *pars basilaris* shapes, forensic anthropologists will gain a better idea of fetus or infant ages since each maturation stage is associated with an age interval. In addition, regarding the WHO definition of viability (more than 22 amenorrhea weeks) and the term of a pregnancy, maturation stages higher than 3 can indicate whether a fetus is viable, and stages 11 and 12 are helpful for marking the term of the pregnancy.

4.5. Maturation and Growth Criterion

In our method, femoral length was chosen as the growth criterion because of its strong relationship with age, and the *pars basilaris* shapes gathered in 19 consensus stages were used to characterize maturation. The grouping of shapes into stages based on consensus shapes with overlaps enable one to obtain a logical continuity of maturation for fetuses and infants while also allowing one to compensate for the low number of individuals of certain age groups.

Growth was defined according to the maturation stages, and we used percentiles, since we sometimes had few individuals per stage. As in any inferential approach based on population sampling and because we are aware that the variability in femur size is not limited to that observed in our samples, which were sometimes of limited size, we widened the range. For this, extreme percentiles were added to either side of the 0 and 100 percentiles. As with growth charts, the use of percentiles allows for growth to be precisely "quantified" with limited statistical bias. Thus, for a given stage, if the length of the femur is below or above the extreme percentiles, growth is considered to be altered.

4.6. The Two Main Advantages of This Coupling Method

The method established in this study makes it possible to analyze the link between the biometric (growth) and physiological (maturation) age of fetuses and infants by coupling the maturation process estimated by means of the *pars basilaris* outline and the growth process estimated by means of the femoral diaphyseal length.

The results obtained from the nonpathological validation sample (B) are encouraging for the fetus and infant age-at-death estimation. We reported coupling in 90.48% of samples, so not only can our method confirm the "overall normality" of this nonpathological sample

(first advantage), but we can also be confident when using a method with femoral length to assess age (second advantage).

Only 4 out of the 42 individuals of sample B showed uncoupling, and they never exceeded a shift of two stages of *pars basilaris* maturation. According to medical reports, these individuals did not have any identified pathological conditions, but in addition to the variability that we tried to include as much as possible in our learning sample (A), several factors can explain uncoupling, such as parity [95–99], parent general height and build [100,101], and the overall progress of the pregnancy, including the exchanges between the fetus and the placenta [100,102–110]. These appear to just have a slightly different variability from our learning sample and confirm that no method can be expected to be 100% reliable due to normal human variability.

4.7. Pathological Uncoupling

As previously mentioned, age estimation from femoral length may be biased since the individual may have had abnormal growth [32], which is not necessarily visible at first sight. This is particularly true when there are no visible bone deformations or malformations such as those which can be seen on fetuses with thanatophoric dysplasia type I-II, osteogenesis imperfecta type IIA, hypophosphatasia, achondrogenesis type IA-II, or diastrophic dysplasia group) [111]. For example, a small stature is found in trisomy 21 fetuses, whose femoral lengths are smaller than normal [112,113] and there are no obvious bone deformations that alert about this pathological state. Additionally, various chromosomal abnormalities or chronic utero-vascular insufficiencies can bias estimations of fetal biometric age [32].

Disease-related bone conditions are not always visible on a skeleton because, for the lesions caused by these conditions to be visible, the individual must be immunologically affected enough to allow disease development yet strong enough to survive it [114]. For example, there are no visible traces on fetal or juvenile human osteological remains of individuals affected by plague, whooping cough, smallpox, measles, scarlet fever, or even osteomyelitis or congenital syphilis, since the disease causes death before any bony stigmas can develop. Thus, childhood disease is not obviously observable from a skeleton, especially when the skeleton is moderately preserved [37].

In our study, uncoupling concerns: localized anomalies, constitutional bone diseases, growth disorders, and cerebral anomalies. Cerebral anomalies are related to size anomalies and malformations: there is one case of cerebral hypotrophy, one case of cerebral gliosis, one case of hydrocephalus, one case of bilateral frontal paraventricular cysts, one case of infection with necrotizing and viro-induced malformative ventriculoencephalitis cytomegalovirus, and one case of agenesis of the corpus callosum associated with microcephaly. Constitutional bone diseases form a heterogeneous group of conditions responsible for insufficient stature or abnormalities in the structure of the bone, whether or not associated with deformities [115]. Among these, uncoupling indicated one case of achondroplasia, one case of Ellis–van Creveld syndrome, one case of Jeune syndrome (or asphyxiating thoracic dysplasia), two cases of thanatophoric dysplasia, one case of femoral-facial syndrome, one VACTERL-type association case, and one case of harlequin ichthyosis.

For all the affected individuals, the femur growth did not match *pars basilaris* maturation. Some authors have further stated that the femoral length is the most suitable biometric parameter for distinguishing bone dysplasias: fetuses with a femur below 30% the mean for gestational age would have achondroplasia; fetuses with a femur between 40% and 60% the mean for gestational age would have thanatophoric dysplasia or type II osteogenesis imperfecta; and fetuses with a femur below 80% the mean for gestational age would be affected by hypochondroplasia, achondroplasia, or type III osteogenesis imperfecta [116].

For uncoupling in individuals with growth disorders, two individuals were found to have diabetes, one macrosomia, four IUGR, and the last one had a twin pregnancy. All these abnormalities or simple variations in growth (twin pregnancy is not necessarily a

pathological pregnancy) could lead to either growth delays or advancements depending on the description of the symptoms, evidence for which can be retrieved with this method.

However, not all individuals in our pathological sample showed systematic uncoupling since the growth disorders associated with each disease depend on several factors such as their origin, their arrival during pregnancy, and their severity. This is the reason why only a few cases were detected. For example, the severity of macrosomia varies according to maternal, pregestational, and gestational diabetes, regardless of association with obesity [117,118]. Macrosomia is also associated with the mother's age (the more advanced, the higher the risk) and parity (the more pregnancies the mother has had, the greater the risk) [118]. Unfortunately, this information cannot be verified since it had not been entered into our database.

Regarding IUGR, a fetus will develop this condition if it cannot achieve its genetic potential for growth due to genetic or external phenomena modifying this potential, or because an abnormality during pregnancy causes growth restriction [119]. Again, the severity of IUGR depends on its cause, the timing of its occurrence during pregnancy, and the duration of the intrauterine aggression [119]. Generally, fetuses with IUGR catch up in terms of their height during the second year of life, often as early as one year [120–122]. A child over 3 years of age who has still not caught up to his height should be taken care of by a pediatrician endocrinologist for in-depth examinations on stature delay, with a view initiating growth hormone treatment from the age of four [121–124]. It should be added that in cases of IUGR, cerebral maturation is generally not affected [125,126].

Additionally, there are variations in growth for multiple pregnancies compared to single pregnancies. For twins, a difference in the mean weight for gestational age is noted from 30 weeks [119]. The differences in growth between twins can be explained by the type of pregnancy; if it is monochorial–biamniotic, the transfusion–transfused syndrome is the first explanation. In bichorium–biamniotic pregnancies, the difference can be explained by a malformation of one of the twins. Placental anomalies and poor fetoplacental exchanges (nutritional, hypoxic, or toxic) can also explain growth anomalies [119].

Finally, the uncoupling of individuals with one or more localized anomalies concern:
Skull anomalies in four cases:

- (1–2) Two microcephaly cases (one was associated with craniosynostosis);
- (3) One ossification defect of the vault with the enlargement of the fontanelles and the presence of Wormian bones in the parietal and occipital region;
- (4) One severe hydrocephalus;

An anomaly of the limbs for one case:

- Anomaly of the femurs with shortening and curving;

An anomaly of the spine for one case:

- A *spina bifida;*

Three cases of polymalformative syndrome:

- One case with arthrogryposis, club feet, clenched hands, 11 pairs of slender ribs and platyspondyly;
- One case with abnormalities of the spine and ribs, as well as retrognathism;
- One case with anomalies of the spine, a short thorax, and a malposition of the four limbs (clenched hands, knees in extension, and club feet).

Finally, the cases of uncoupling highlighted by our method suggest that when maturation and growth do not match, experts must be prepared for a possible anomaly or variation in growth that risks biasing the age as estimated from femoral length.

Thus, the proposed method should be used in forensic anthropology for age estimation to verify whether growth has been altered by possible pathological conditions. This appears to be crucial in forensic contexts, where age estimation should be as accurate as possible to assess viability, set at 22 weeks of amenorrhea or a weight of 500 g according to WHO

recommendations, to determine whether an individual came to term and to provide an unbiased age-at-death for police investigations.

To improve this method in the future, it would be of interest to include more healthy individuals to reduce the age range for some stages in order to provide greater precision in determining the consensus shape. The inclusion of samples from various origins would also allow the method to be used in different populations, and it could also be used in a clinical setting for screening for abnormal growth.

5. Conclusions

This study was focused on characterizing the link between maturation and growth by analyzing bone shape and biometry. The use of geometric morphometric methods and elliptical Fourier analysis enabled us to precisely quantify the *pars basilaris* shape changes from 16 fetal weeks to approximately one and a half years (17.7 months) in an unprecedented way.

By considering the coupling between the maturation and growth process, it is possible to detect potential anomalies or variations in growth. It is important to remember that it is difficult to macroscopically detect bone anomalies that could alert one to this possible variation and that the application of age-at-death estimation methods can be biased since they were established from reference populations with normal development but that the targeted individuals do not necessarily meet this condition.

In cases of uncoupling, experts should be warned that living conditions have altered the development of a young individual and that the age-at-death estimation based on long bone biometry may be biased. In a forensic context, the detection of uncoupling must lead an expert to be careful in their conclusions regarding the age determined for a young juvenile.

Author Contributions: Conceptualization, methodology, validation, and visualization: M.N. and P.A. Software, formal analysis, investigation, data curation, and writing—original draft: M.N. Writing—review and editing, supervision, and project administration: P.A. Resources: K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Public Hospital of Marseille (AP-HM) (code CIL/AP-HM 2017-41) and The National Commission on Informatics and Liberties (CNIL) (code 2071641 v 0 registered on the 27 June 2017).

Informed Consent Statement: Informed consent was obtained from the mothers of all young subjects involved in the study.

Data Availability Statement: As mentioned in the informed consent signed by the mothers, the data for individuals will remain strictly confidential.

Acknowledgments: The authors wish to thank Clémence Delteil, Louise Corron, Emmanuelle Lesieur, and Floriane Remy for the co-establishment of the immature database of the UMR 7268 ADES (AMU, CNRS, EFS).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CA	Cerebral Anomalies
CBD	Constitutional Bone Diseases
CHRA	Chromosomal Anomalies
CT scan	Computerized Tomography Scan
DICOM	Digital Imaging and Communications in Medicine
EFA	Elliptical Fourier Analysis
GD	Growth Disorders

GPA	Generalized Procrustes Analysis
IUGR	Intrauterine Growth Retardation
LA	Localized Anomalies
MANOVA	Multivariate Analysis Of Variance
MRI	Magnetic Resonance Imaging
VACTERL	Vertebral, Anal, Cardiac, Tracheal, Esophageal, Renal, and Limb
WHO	World Health Organization

References

1. Ritz-Timme, S.; Cattaneo, C.; Collins, M.J.; Waite, E.R.; Schutz, H.W.; Kaatsch, H.J.; Borrman, H.I.M. Age Estimation: The State of the Art in Relation to the Specific Demands of Forensic Practise. *Int. J. Legal Med.* **2000**, *113*, 129–136. [CrossRef]
2. Adalian, P. *Evaluation Multiparamétrique de la Croissance Foetale: Application à la Détermination de L'âge et du Sexe*; Université de la Méditerranée-Aix-Marseille 2: Marseille, France, 2001. Available online: http://www.theses.fr/2001AIX20692 (accessed on 5 April 2018).
3. Adalian, P.; Piercecchi-Marti, M.D.; Bourliere-Najean, B.; Panuel, M.; Leonetti, G.; Dutour, O. New foetal age determination formula. *Comptes Rendus Biol.* **2002**, *325*, 261–269. [CrossRef]
4. AlQahtani, S.J.; Hector, M.P.; Liversidge, H.M. Brief Communication: The London Atlas of Human Tooth Development and Eruption. *Am. J. Phys. Anthropol.* **2010**, *142*, 481–490. [CrossRef]
5. Balthazard, V.; Dervieux, X. Etudes Anthropométriques Sur Le Foetus Humain. *Ann. Med. Leg.* **1921**, *1*, 37–42.
6. Cardoso, H.F.V.; Gomes, J.; Campanacho, V.; Marinho, L. Age Estimation of Immature Human Skeletal Remains Using the Post-Natal Development of the Occipital Bone. *Int. J. Legal Med.* **2013**, *127*, 997–1004. [CrossRef]
7. Cardoso, H.F.V.; Abrantes, J.; Humphrey, L.T. Age Estimation of Immature Human Skeletal Remains from the Diaphyseal Length of the Long Bones in the Postnatal Period. *Int. J. Legal Med.* **2014**, *128*, 809–824. [CrossRef]
8. Carneiro, C.; Curate, F.; Borralho, P.; Cunha, E. Radiographic Fetal Osteometry: Approach on Age Estimation for the Portuguese Population. *Forensic Sci. Int.* **2013**, *231*, 397.e1. [CrossRef]
9. Carneiro, C.; Curate, F.; Cunha, E. A Method for Estimating Gestational Age of Fetal Remains Based on Long Bone Lengths. *Int. J. Legal Med.* **2016**, *130*, 1333–1341. [CrossRef]
10. Chavez-Martinez, P.; Ortega-Palma, A.; Castrejón-Caballero, J.L.; Arteaga-Martínez, M. Equations to Estimate Fetal Age at the Moment of Death in the Mexican Population. *Forensic Sci. Int.* **2016**, *266*, 587.e1–587.e10. [CrossRef]
11. Fazekas, I.G.; Kósa, F. *Forensic Fetal Osteology*; Akadémiai Kiadó: Budapest, Hongrie, 1978; ISBN 978-963-05-1491-0.
12. Hadlock, F.; Harrist, R.; Deter, R.; Park, S. Fetal Femur Length as a Predictor of Menstrual Age: Sonographically Measured. *Am. J. Roentgenol.* **1982**, *138*, 875–878. [CrossRef]
13. Kurmanavicius, J.; Wright, E.M.; Royston, P.; Zimmermann, R.; Huch, R.; Huch, A.; Wisser, J. Fetal Ultrasound Biometry: 2. Abdomen and Femur Length Reference Values. *BJOG Int. J. Obstet. Gynaecol.* **1999**, *106*, 136–143. [CrossRef]
14. Maresh, M.M. Linear Growth of Long Bones of Extremities from Infancy through Adolescence: Continuing Studies. *AMA Am. J. Dis. Child.* **1955**, *89*, 725–742. [CrossRef]
15. Olivares, J.I.; Aguilera, I.A. Proposal of New Regression Formulae for the Estimation of Age in Infant Skeletal Remains from the Metric Study of the Pars Basilaris. *Int. J. Legal Med.* **2017**, *131*, 781–788. [CrossRef]
16. Olivier, G.; Pineau, H. Nouvelle Détermination de La Taille Foetale d'après Les Longueurs Diaphysaires Des Os Longs. *Ann. Med. Leg.* **1960**, *40*, 141–144.
17. Scheuer, L.; Musgrave, J.; Evans, S. The Estimation of Late Fetal and Perinatal Age from Limb Bone Length by Linear and Logarithmic Regression. *Ann. Hum. Biol.* **1980**, *7*, 257–265. [CrossRef]
18. Scheuer, L.; MacLaughlin-Black, S. Age Estimation from the Pars Basilaris of the Fetal and Juvenile Occipital Bone. *Int. J. Osteoarchaeol.* **1994**, *4*, 377–380. [CrossRef]
19. Simon, D.; Bhowmick, B.; Simon, M.; Varghese, D.; Varghese, A. Fetal Growth and Age Estimation Based on Osteometry on the Diaphysis of Human Fetal Fibula. *Int. J. Innov. Res. Dev.* **2020**, *9*, 49–57. [CrossRef]
20. Simon, D.; Varghese, A.; Bhowmick, B.; Simon, M.; Varghese, D. Human Fetal Growth and Age Estimation Based on Osteometry on the Diaphysis of Fetal Tibia. *Int. J. Innov. Res. Dev.* **2020**, *9*, 40–48. [CrossRef]
21. Feess-Higgins, A.; Larroche, J.-C. *Development of the Human Fetal Brain: Anatomical Atlas*; Editions Inserm: Paris, France, 1987; ISBN 978-2-85598-337-0.
22. Piercecchi-Marti, M.D.; Adalian, P.; Pelissier-Alicot, A.-L.; Cianfarani, F.; Leonetti, G. Estimation de l'âge Foetal à Partir de l'examen Macroscopique et Histologique Des Viscères. *J. Médecine Légale Droit Médical* **2000**, *43*, 545–552. Available online: http://cat.inist.fr/?aModele=afficheN&cpsidt=972452 (accessed on 18 May 2015).
23. Scheuer, L.; Black, S.; Cunningham, C. *Developmental Juvenile Osteology*; Academic Press: Cambridge, MA, USA, 2000; ISBN 978-0-08-053038-3.
24. Captier, G.; Boë, L.-J.; Badin, P.; Guihard-Costa, A.-M.; Canovas, F.; Larroche, J.-C. Modèles Géométriques de Croissance Du Cerveau, Cervelet, Tronc Cérébral et Modification Des Angles de La Base Du Crâne Au Cours de La Période Fœtale. *Morphologie* **2013**, *97*, 38–47. [CrossRef]
25. Ford, E.H. The Growth of the Foetal Skull. *J. Anat.* **1956**, *90*, 63–72. [PubMed]

26. Grossman, J.W.; Zuckerman, S. An X-ray Study of Growth Changes in the Base of the Skull. *Am. J. Phys. Anthropol.* **1955**, *13*, 515–519. [CrossRef] [PubMed]
27. Jeffery, N.; Spoor, F. Brain Size and the Human Cranial Base: A Prenatal Perspective. *Am. J. Phys. Anthropol.* **2002**, *118*, 324–340. [CrossRef] [PubMed]
28. Lieberman, D.E.; Pearson, O.M.; Mowbray, K.M. Basicranial Influence on Overall Cranial Shape. *J. Hum. Evol.* **2000**, *38*, 291–315. [CrossRef]
29. Mandarim de lacerda, C.; Alves, M. Growth of the Cranial Bones in Human Fetuses (2nd and 3rd Trimesters). *Surg. Radiol. Anat.* **1992**, *14*, 125–129. [CrossRef]
30. Morimoto, N.; Ogihara, N.; Katayama, K.; Shiota, K. Three-Dimensional Ontogenetic Shape Changes in the Human Cranium during the Fetal Period. *J. Anat.* **2008**, *212*, 627–635. [CrossRef]
31. Trenouth, M.J. Shape Changes during Human Fetal Craniofacial Growth. *J. Anat.* **1984**, *139*, 639–651.
32. Sherwood, R.; Meindl, R.S.; Robinson, H.B.; May, R.L. Fetal Age: Methods of Estimation and Effects of Pathology. *Am. J. Phys. Anthropol.* **2000**, *113*, 305–315. [CrossRef]
33. Delteil, C.; Lesieur, E.; Tuchtan, L.; Álvarez, A.C.; Chaumoitre, K.; Saliba, B.; Adalian, P.; Piercecchi-Marti, M. Study of the Growth and Shape of the Brain and Cranial Base during the First Two Years of Life. *Morphol. Bull. Assoc. Anat.* **2020**, *105*, 45–53. [CrossRef]
34. Delteil, C.; Mopin, C.; Tuchtan, L.; Alvarez, A.; Piercecchi, M.-D.; Adalian, P. Ontogeny of Cranial Base during the First Two Years of Life. *Morphologie* **2021**, *in press*. [CrossRef] [PubMed]
35. Sastre, J.; Tardivo, D.; Khemira, A.; Chaumoitre, K.; Panuel, M.; Piercecchi-Marti, M.D.; Leonetti, G.; Adalian, P. 3D CT Scan Study of Fetal Cranial Base: Interests of Occipital Bone Measurements in Age Estimation. *Odonto-Stomatol. Trop. Trop. Dent. J.* **2011**, *34*, 17–25.
36. Sastre, J. *Etude Tridimensionnelle de la Partie Postérieure de la Base du Crâne chez le Fœtus: Applications à la Détermination de L'âge en Anthropologie Médico-Légale*; Université de la Méditerranée-Aix Marseille 2: Marseille, France, 2011.
37. Lewis, M.E. *The Bioarchaeology of Children: Perspectives from Biological and Forensic Anthropology*; Cambridge University Press: New York, NY, USA, 2006; ISBN 978-1-139-45968-6.
38. Coleman, M.N.; Colbert, M.W. Technical Note: CT Thresholding Protocols for Taking Measurements on Three-Dimensional Models. *Am. J. Phys. Anthropol.* **2007**, *133*, 723–725. [CrossRef] [PubMed]
39. Spoor, F.; Zonneveld, F.W.; Macho, G.A. Linear Measurements of Cortical Bone and Dental Enamel by Computed Tomography: Applications and Problems. *Am. J. Phys. Anthropol.* **1993**, *91*, 469–484. [CrossRef] [PubMed]
40. Adalian, P.; Piercecchi-Marti, M.D.; Bourliere-Najean, B.; Panuel, M.; Fredouille, C.; Dutour, O.; Leonetti, G. Postmortem Assessment of Fetal Diaphyseal Femoral Length: Validation of a Radiographic Methodology. *J. Forensic Sci.* **2001**, *46*, 215–219. [CrossRef]
41. Bareggi, R.; Grill, V.; Zweyer, M.; Sandrucci, M.A.; Martelli, A.M.; Narducci, P.; Forabosco, A. On the Assessment of the Growth Patterns in Human Fetal Limbs: Longitudinal Measurements and Allometric Analysis. *Early Hum. Dev.* **1996**, *45*, 11–25. [CrossRef]
42. Haj Salem, N.; Dedouit, F.; Maret, D.; Adalian, P.; Rougé, D.; Piercecchi, M.-D.; Assaoui, A.; Chadly, A.; Norbert, T. Estimation de l'âge foetal par étude scanographique des longueurs fémorales et tibiales: Comparaison entre les reconstructions en deux et en trois dimensions. *Rev. Médecine Légale* **2010**, *1*, 71–76. [CrossRef]
43. Niel, M.; Chaumoître, K.; Corny, J.; Lalys, L.; Adalian, P. Maturation of the Human Foetal Basioccipital: Quantifying Shape Changes in Second and Third Trimesters Using Elliptic Fourier Analysis. *J. Anat.* **2019**, *235*, 34–44. [CrossRef]
44. Bookstein, F.L. *Morphometric Tools for Landmark Data: Geometry and Biology*; Cambridge University Press: New York, NY, USA, 1991.
45. Rohlf, F.J. *TpsDig2: Digitize Landmarks and Outlines*; Department of Ecology, Evolution and Anthropology, State University of New York at Stony Brook: New York, NY, USA, 2013. Available online: http://life.bio.sunysb.edu/morph/soft-dataacq.html (accessed on 13 April 2021).
46. Friess, M.; Baylac, M. Exploring Artificial Cranial Deformation Using Elliptic Fourier Analysis of Procrustes Aligned Outlines. *Am. J. Phys. Anthropol.* **2003**, *122*, 11–22. [CrossRef]
47. Gower, J.C. Generalized Procrustes Analysis. *Psychometrika* **1975**, *40*, 33–51. [CrossRef]
48. Rohlf, F.J. Statistical Power Comparisons among Alternative Morphometric Methods. *Am. J. Phys. Anthropol.* **2000**, *111*, 463–478. [CrossRef]
49. Rohlf, F.J.; Slice, D. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst. Biol.* **1990**, *39*, 40–59. [CrossRef]
50. Shannon, C.E. Communication in the Presence of Noise. *Proc. IRE* **1949**, *37*, 10–21. [CrossRef]
51. Claude, J. Log-Shape Ratios, Procrustes Superimposition, Elliptic Fourier Analysis: Three Worked Examples in R. *Hystrix-Ital. J. Mammal.* **2013**, *24*, 94–102. [CrossRef]
52. Claude, J.; Paradis, E.; Tong, H.; Auffray, J.C. A Geometric Morphometric Assessment of the Effects of Environment and Cladogenesis on the Evolution of the Turtle Shell. *Biol. J. Linn. Soc.* **2003**, *79*, 485–501. [CrossRef]
53. Debat, V.; Cornette, R.; Korol, A.B.; Nevo, E.; Soulet, D.; David, J.R. Multidimensional Analysis of Drosophila Wing Variation in Evolution Canyon. *J. Genet.* **2008**, *87*, 407–419. [CrossRef]
54. Goodall, C. Procrustes Methods in the Statistical Analysis of Shape. *J. R. Stat. Soc. Ser. B Methodol.* **1991**, *53*, 285–339. [CrossRef]

55. Klingenberg, C.P.; McIntyre, G.S. Geometric Morphometrics of Developmental Instability: Analyzing Patterns of Fluctuating Asymmetry with Procrustes Methods. *Evolution* **1998**, *52*, 1363–1375. [CrossRef]
56. Claude, J. *Morphometrics with R*; Springer Science & Business Media: New York, NY, USA, 2008; ISBN 978-0-387-77790-0.
57. Monti, L.; Baylac, M.; Lalanne-Cassou, B. Elliptic Fourier Analysis of the Form of Genitalia in Two Spodoptera Species and Their Hybrids (Lepidoptera: Noctuidae). *Biol. J. Linn. Soc.* **2001**, *72*, 391–400. [CrossRef]
58. Rohlf, F.J.; Archie, J.W. A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae). *Syst. Biol.* **1984**, *33*, 302–317. [CrossRef]
59. Anderson, M. A New Method for Non-Parametric Multivariate Analysis of Variance. *Austral Ecol.* **2001**, *26*, 32–46. [CrossRef]
60. Mitteroecker, P.; Gunz, P. Advances in Geometric Morphometrics. *Evol. Biol.* **2009**, *36*, 235–247. [CrossRef]
61. Zelditch, M.L.; Swiderski, D.L.; Sheets, H.D. *Geometric Morphometrics for Biologists, Second Edition: A Primer*, 2nd ed.; Academic Press: Amsterdam, The Netherlands, 2012; ISBN 978-0-12-386903-6.
62. Bonhomme, V.; Claude, J. Momocs: Morphometrics Using R; 2017. Available online: https://CRAN.R-project.org/package=Momocs (accessed on 1 March 2018).
63. Schlager, S.; Jefferis, G. *Morpho: Calculations and Visualisations Related to Geometric Morphometrics*; 2017. Available online: https://CRAN.R-project.org/package=Morpho (accessed on 1 March 2018).
64. Adams, D.; Collyer, M.; Kaliontzopoulou, A.; Sherratt, E. *Geomorph: Geometric Morphometric Analyses of 2D/3D Landmark Data*; 2017. Available online: https://cran.r-project.org/package=geomorph (accessed on 1 March 2018).
65. Fox, J.; Weisberg, S.; Price, B.; Adler, D.; Bates, D.; Baud-Bovy, G.; Bolker, B.; Ellison, S.; Firth, D.; Friendly, M.; et al. *Car: Companion to Applied Regression*; 2018. Available online: https://CRAN.R-project.org/package=car (accessed on 10 April 2018).
66. Zhao, J.H.; Hornik, K.; Ripley, B. *Gap: Genetic Analysis Package*; 2018. Available online: https://CRAN.R-project.org/package=gap (accessed on 10 April 2018).
67. Cordoliani, Y.-S.; Foehrenbach, H.; Dion, A.-M.; Lahutte, M. Risques liés à l'exposition d'une grossesse aux rayonnements ionisants. *J. Radiol.* **2005**, *86*, 601–606. [CrossRef]
68. Chen, S.Y.; Lestrel, P.E.; Kerr, W.J.; McColl, J.H. Describing Shape Changes in the Human Mandible Using Elliptical Fourier Functions. *Eur. J. Orthod.* **2000**, *22*, 205–216. [CrossRef] [PubMed]
69. Corina Vera, M.; Laura Ponssa, M.; Abdala, V. Further Data on Sesamoid Identity from Two Anuran Species. *Anat. Rec.-Adv. Integr. Anat. Evol. Biol.* **2015**, *298*, 1376–1394. [CrossRef] [PubMed]
70. Corny, J.; Detroit, F. Technical Note: Anatomic Identification of Isolated Modern Human Molars: Testing Procrustes Aligned Outlines as a Standardization Procedure for Elliptic Fourier Analysis. *Am. J. Phys. Anthropol.* **2014**, *153*, 314–322. [CrossRef] [PubMed]
71. Daegling, D.J.; Jungers, W.L. Elliptical Fourier Analysis of Symphyseal Shape in Great Ape Mandibles. *J. Hum. Evol.* **2000**, *39*, 107–122. [CrossRef]
72. Ferson, S.; Rohlf, F.J.; Koehn, R.K. Measuring Shape Variation of Two-Dimensional Outlines. *Syst. Biol.* **1985**, *34*, 59–68. [CrossRef]
73. Lestrel, P.E. *Fourier Descriptors and Their Applications in Biology*; Cambridge University Press: New York, NY, USA, 2008; ISBN 978-0-521-05573-4.
74. Lestrel, P.E.; Bodt, A.; Swindler, D.R. Longitudinal Study of Cranial Base Shape Changes in Macaca Nemestrina. *Am. J. Phys. Anthropol.* **1993**, *91*, 117–129. [CrossRef]
75. Mokrane, F.-Z.; Dedouit, F.; Gellee, S.; Sans, N.; Rousseau, H.; Rouge, D.; Telmon, N. Sexual Dimorphism of the Fetal Ilium: A 3D Geometric Morphometric Approach with Multislice Computed Tomography. *J. Forensic Sci.* **2013**, *58*, 851–858. [CrossRef]
76. Caple, J.; Byrd, J.; Stephan, C.N. Elliptical Fourier Analysis: Fundamentals, Applications, and Value for Forensic Anthropology. *Int. J. Legal Med.* **2017**, *131*, 1675–1690. [CrossRef]
77. Kuhl, F.P.; Giardina, C.R. Elliptic Fourier Features of a Closed Contour. *Comput. Graph. Image Process.* **1982**, *18*, 236–258. [CrossRef]
78. Jeffery, N.; Spoor, F. Ossification and Midline Shape Changes of the Human Fetal Cranial Base. *Am. J. Phys. Anthropol.* **2004**, *123*, 78–90. [CrossRef]
79. Macklin, C.C. Preliminary Note on the Skull of a Human Fetus of 43 MM. Greatest Length. *Anat. Rec.* **1921**, *22*, 251–265. [CrossRef]
80. Nemzek, W.R.; Brodie, H.A.; Hecht, S.T.; Chong, B.W.; Babcook, C.J.; Seibert, J.A. MR, CT, and Plain Film Imaging of the Developing Skull Base in Fetal Specimens. *Am. J. Neuroradiol.* **2000**, *21*, 1699–1706. [PubMed]
81. Noback, C.R. The Developmental Anatomy of the Human Osseous Skeleton during the Embryonic, Fetal and Circumnatal Periods. *Anat. Rec.* **1944**, *88*, 91–125. [CrossRef]
82. Noback, C.R.; Robertson, G.G. Sequences of Appearance of Ossification Centers in the Human Skeleton during the First Five Prenatal Months. *Am. J. Anat.* **1951**, *89*, 1–28. [CrossRef]
83. Redfield, A. A New Aid to Aging Immature Skeletons: Development of the Occipital Bone. *Am. J. Phys. Anthropol.* **1970**, *33*, 207–220. [CrossRef] [PubMed]
84. Schaefer, M.; Black, S.; Scheuer, L. *Juvenile Osteology: A Laboratory and Field Manual*; Elsevier: Amsterdam, The Netherlands; Academic Press: Burlington, MA, USA, 2009; ISBN 978-0-12-374635-1.
85. Scheuer, L.; Black, S. *The Juvenile Skeleton*; Elsevier Academic Press: London, UK, 2004; ISBN 978-0-08-047459-5.
86. Slice, D.E. Geometric Morphometrics. *Annu. Rev. Anthropol.* **2007**, *36*, 261–281. [CrossRef]
87. Needham, A.E. The Form-Transformation of the Abdomen of the Female Peacrab, Pinnotheres Pisum Leach. *Proc. R. Soc. Lond. B Biol. Sci.* **1950**, *137*, 115–136.

88. Herlin, C.; Largey, A.; deMattei, C.; Daures, J.P.; Bigorre, M.; Captier, G. Modeling of the Human Fetal Skull Base Growth: Interest in New Volumetrics Morphometric Tools. *Early Hum. Dev.* **2011**, *87*, 239–245. [CrossRef]
89. Jeffery, N. A High-Resolution MRI Study of Linear Growth of the Human Fetal Skull Base. *Neuroradiology* **2002**, *44*, 358–366. [CrossRef]
90. Schmeling, A.; Geserick, G.; Reisinger, W.; Olze, A. Age Estimation. *Forensic Sci. Int.* **2007**, *165*, 178–181. [CrossRef] [PubMed]
91. Tocheri, M.W.; Molto, J.E. Aging Fetal and Juvenile Skeletons from Roman Period Egypt Using Basiocciput Osteometrics. *Int. J. Osteoarchaeol.* **2002**, *12*, 356–363. [CrossRef]
92. Hill, L.M.; Guzick, D.; Hixson, J.; Peterson, C.S.; Rivello, D.M. Composite Assessment of Gestational Age: A Comparison of Institutionally Derived and Published Regression Equations. *Am. J. Obstet. Gynecol.* **1992**, *166*, 551–555. [CrossRef]
93. Piercecchi-Marti, M.D.; Adalian, P.; Liprandi, A.; Figarella-Branger, D.; Dutour, O.; Leonetti, G. Fetal Visceral Maturation: A Useful Contribution to Gestational Age Estimation in Human Fetuses. *J. Forensic Sci.* **2004**, *49*, 912–917. [CrossRef] [PubMed]
94. Kyrkanides, S.; Kjaer, I.; Fischerhansen, B. Development of the Basilar Part of the Occipital Bone in Normal Human Fetuses. *J. Craniofac. Genet. Dev. Biol.* **1993**, *13*, 184–192.
95. Gaillard, R.; Rurangirwa, A.; Williams, M.A.; Hofman, A.; Mackenbach, J.P.; Franco, O.H.; Steegers, E.A.P.; Jaddoe, V.W.V. Maternal Parity, Fetal and Childhood Growth, and Cardiometabolic Risk Factors. *Hypertension* **2014**, *64*, 266–274. [CrossRef]
96. Hinkle, S.N.; Albert, P.S.; Mendola, P.; Sjaarda, L.A.; Yeung, E.; Boghossian, N.S.; Laughon, S.K. The Association between Parity and Birthweight in a Longitudinal Consecutive Pregnancy Cohort. *Paediatr. Perinat. Epidemiol.* **2014**, *28*, 106–115. [CrossRef]
97. Ong, K.K.L.; Preece, M.A.; Emmett, P.M.; Ahmed, M.L.; Dunger, D.B.; ALSPAC Study Team. Size at Birth and Early Childhood Growth in Relation to Maternal Smoking, Parity and Infant Breast-Feeding: Longitudinal Birth Cohort Study and Analysis. *Pediatr. Res.* **2002**, *52*, 863–867. [CrossRef]
98. Prentice, A.; Cole, T.; Whitehead, R.G. Impaired Growth in Infants Born to Mothers of Very High Parity. *Hum. Nutr. Clin. Nutr.* **1987**, *41*, 319–325.
99. Wilcox, M.A.; Chang, A.M.; Johnson, I.R. The Effects of Parity on Birthweight Using Successive Pregnancies. *Acta Obstet. Gynecol. Scand.* **1996**, *75*, 459–463. [CrossRef] [PubMed]
100. Lejarraga, H. Growth in Infancy and Childhood: A Pediatric Approach. In *Human Growth and Development*; Academic Press: Cambridge, MA, USA, 2012; pp. 21–44.
101. Smith, D.W.; Truog, W.; Rogers, J.E.; Greitzer, L.J.; Skinner, A.L.; McCann, J.J.; Harvey, M.A. Shifting Linear Growth during Infancy: Illustration of Genetic Factors in Growth from Fetal Life through Infancy. *J. Pediatr.* **1976**, *89*, 225–230. [CrossRef]
102. Cameron, N.; Bogin, B. *Human Growth and Development*; Academic Press: Cambridge, MA, USA, 2012; ISBN 978-0-12-383882-7.
103. Kawakita, A.; Sato, K.; Makino, H.; Ikegami, H.; Takayama, S.; Toyama, Y.; Umezawa, A. Nicotine Acts on Growth Plate Chondrocytes to Delay Skeletal Growth through the A7 Neuronal Nicotinic Acetylcholine Receptor. *PLoS ONE* **2008**, *3*. [CrossRef] [PubMed]
104. Koren, G. Measurement of Drugs in Neonatal Hair; a Window to Fetal Exposure. *Forensic Sci. Int.* **1995**, *70*, 77–82. [CrossRef]
105. Lejeune, C. Syndrome d'alcoolisation fœtale. *Devenir* **2001**, *13*, 77–94. [CrossRef]
106. Merialdi, M.; Caulfield, L.E.; Zavaleta, N.; Figueroa, A.; Costigan, K.A.; Dominici, F.; Dipietro, J.A. Randomized Controlled Trial of Prenatal Zinc Supplementation and Fetal Bone Growth. *Am. J. Clin. Nutr.* **2004**, *79*, 826–830. [CrossRef]
107. Samper, M.P.; Jiménez-Muro, A.; Nerín, I.; Marqueta, A.; Ventura, P.; Rodríguez, G. Maternal Active Smoking and Newborn Body Composition. *Early Hum. Dev.* **2012**, *88*, 141–145. [CrossRef]
108. Simpson, M.E.; Duggal, S.; Keiver, K. Prenatal Ethanol Exposure Has Differential Effects on Fetal Growth and Skeletal Ossification. *Bone* **2005**, *36*, 521–532. [CrossRef]
109. Sobrian, S.K.; Vaughn, V.T.; Ashe, W.K.; Markovic, B.; Djuric, V.; Jankovic, B.D. Gestational Exposure to Loud Noise Alters the Development and Postnatal Responsiveness of Humoral and Cellular Components of the Immune System in Offspring. *Environ. Res.* **1997**, *73*, 227–241. [CrossRef]
110. Tabet, C.; Dupuis-Gauthier, C.; Schmidt, P.; Maerten-Lesot, B.; Porez, S.; Delion, P.; Soulé, M. Maltraitance à fœtus: Comment comprendre pour prévenir. *Devenir* **2009**, *21*, 205–244. [CrossRef]
111. *Fetal and Neonatal Pathology*, 4th ed; Keeling, J.W.; Khong, T.Y. (Eds.) Springer: London, UK, 2007; ISBN 978-1-84628-743-5.
112. Benacerraf, B.; Gelman, R.; Frigoletto, F. Sonographic Identification of 2nd-Trimester Fetuses with Downs-Syndrome. *N. Engl. J. Med.* **1987**, *317*, 1371–1376. [CrossRef] [PubMed]
113. Fitzsimmons, J.; Droste, S.; Shepard, T.; Pascoemason, J.; Chinn, A.; Mack, L. Long-Bone Growth in Fetuses with down Syndrome. *Am. J. Obstet. Gynecol.* **1989**, *161*, 1174–1177. [CrossRef]
114. Ortner, D.J. *Identification of Pathological Conditions in Human Skeletal Remains*, 2nd ed.; Academic Press: San Diego, CA, USA, 2003; 664p, ISBN 978-0-12-528628-2.
115. Le Merrer, M.; Lambot, K. Maladies Osseuses Constitutionnelles. *Radiol. Imag. MÉDICALE Musculosquelettique-Neurol.-Maxillofac.* **2012**, *7*, 1–19. [CrossRef]
116. Goncalves, L.; Jeanty, P.J. Fetal Biometry of Skeletal Dysplasias. *J Ultrasound Med* **1994**, *13*, 977–985. [CrossRef] [PubMed]
117. Hauguel-de Mouzon, S.; Lepercq, J. Physiopathologie de La Croissance Du Fœtus de Mère Diabétique. *Médecine Thérapeutique Endocrinol.* **2001**, *3*, 49–52.
118. Touhami Elouazzani, F.; Kabiri, M.; Karboubi, L.; Keswati, J.; Mrabet, M.; Barkat, A. La Macrosomie: À Propos de 255 °C as. *J. Pédiatrie Puériculture* **2012**, *25*, 97–101. [CrossRef]

119. Fournié, A.; Kessler, S.; Biquard, F.; Parant, O.; Connan, L. Hypotrophie, Retard de Croissance Intra-Utérin, Souffrance Fœtale Chronique. *EMC-Gynécologie-Obs.* **2004**, *1*, 97–126. [CrossRef]
120. Egounlety, F. *Suivi Entre 1 et 3 ans D'enfants nés Petits Pour L'âge Gestationnel ("RCIU"): Étude Descriptive Multicentrique de 2 Populations Issues du Réseau Pédiatrique Sud et Ouest Francilien, les Grands Prématurés RICU et les Nouveau-nés de Plus de 33 SA RICU A.*; Université de Lorraine: Nancy, France, 2016.
121. Lee, P.A.; Chernausek, S.D.; Hokken-Koelega, A.C.S.; Czernichow, P. International Small for Gestational Age Advisory Board Consensus Development Conference Statement: Management of Short Children Born Small for Gestational Age, April 24–October 1, 2001. *Pediatrics* **2003**, *111*, 1253–1261. [CrossRef]
122. Urs Zumsteg, B. RCIU et retard de croissance-indication pour l'hormone de croissance. *Paediatrica* **2011**, *22*, 12–13.
123. Gascoin, G.; Flamant, C. Conséquences à long terme des enfants nés dans un contexte de retard de croissance intra-utérin et/ou petits pour l'âge gestationnel. *J. Gynécologie Obs. Biol. Reprod.* **2013**, *42*, 911–920. [CrossRef]
124. Léger, J. L'enfant Né Petit Pour l'âge Gestationnel: Sa Croissance, Son Devenir. *Médecine Thérapeutique Pédiatrie* **2006**, *9*, 242–250.
125. Larroche, J.C. *Developmental Pathology of the Neonate*; Excerpta Medica: Paris, France, 1977; ISBN 978-90-219-2107-5.
126. Singer, D.; Sung, C.; Wigglesworth, J. Fetal Growth and Maturation: With Standards for Body and Organ Development. In *Textbook of Fetal and Perinatal Pathology*; Blackwell Scientific: Cambridge, MA, USA, 1991; pp. 11–47.

biology

MDPI

Article

Identifying Blunt Force Traumatic Injury on Thermally Altered Remains: A Pilot Study Using *Sus scrofa*

Kamryn Keys and Ann H. Ross *

Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA; kkeys@ncsu.edu
* Correspondence: ahross@ncsu.edu

Simple Summary: Human remains are often burned in an effort to conceal the identity of the victim and/or obscure traumatic injuries related to the death event. Thermal exposure can produce artifacts resembling trauma and disguise preexisting trauma. However, there is a paucity of experimental studies with varied results addressing the differentiation of thermally induced artifacts from traumatic signatures. To address this gap in the literature, we conducted a small-scale study using domestic pigs as correlates to test the impact of thermal alteration on blunt force trauma to the cranium. Two tools (e.g., hammer and crowbar) were utilized to manually inflict injuries on the human analogs before controlled burning in an outdoor environment. The results of this experiment demonstrated that the most diagnostic variable to differentiate thermally induced alternations from blunt force fractures was fracture pattern.

Abstract: In forensic scenarios involving homicide, human remains are often exposed to fire as a means of disposal and/or obscuring identity. Burning human remains can result in the concealment of traumatic injury, the creation of artifacts resembling injury, or the destruction of preexisting trauma. Since fire exposure can greatly influence trauma preservation, methods to differentiate trauma signatures from burning artifacts are necessary to conduct forensic analyses. Specifically, in the field of forensic anthropology, criteria to distinguish trauma from fire signatures on bone is inconsistent and sparse. This study aims to supplement current forensic anthropological literature by identifying criteria found to be the most diagnostic of fire damage or blunt force trauma. Using the skulls of 11 adult pigs (*Sus scrofa*), blunt force trauma was manually produced using a crowbar and flat-faced hammer. Three specimens received no impacts and were utilized as controls. All skulls were relocated to an outdoor, open-air fire where they were burned until a calcined state was achieved across all samples. Results from this experiment found that blunt force trauma signatures remained after burning and were identifiable in all samples where reassociation of fragments was possible. This study concludes that distinct patterns attributed to thermal fractures and blunt force fractures are identifiable, allowing for diagnostic criteria to be narrowed down for future analyses.

Keywords: forensic anthropology; forensic science; blunt force trauma; thermal alteration; thermal fractures

Citation: Keys, K.; Ross, A.H. Identifying Blunt Force Traumatic Injury on Thermally Altered Remains: A Pilot Study Using *Sus scrofa*. *Biology* **2022**, *11*, 87. https://doi.org/10.3390/biology11010087

Academic Editor: Andrés Moya

Received: 21 November 2021
Accepted: 22 December 2021
Published: 6 January 2022

1. Introduction

Trauma interpretation is arguably one of the most valuable services a forensic anthropologist can perform to assist criminal investigative proceedings. This is evidenced by the consistent theme of trauma-focused research in the forensic anthropology literature, spanning several decades [1–12]. Special interest is, in part, due to the fact that biomechanical signatures of skeletal trauma are not fully understood [13–16]. As such, considerable research replicating traumatic force has been produced to document the resultant characteristics seen on bone [1–11,13]. Efforts to identify the source of trauma are only half of the assessment, as it is imperative for the timing of the injury to be established as well. To interpret injury timing, characteristics of the defect are noted concerning the reaction

of the immediate surrounding bone [8,9,13–16]. Following visual observation of trauma, it is assigned to an ante-, peri-, or postmortem temporal context [3,4,8]. By establishing the timing of the defect, the anthropologist can provide insight into whether the injury potentially contributed to the death event [3,4,8,13–16].

Although the characteristics of blunt force impact are the focus of many studies [1–4,6–8,11,13,16], there is still much ambiguity surrounding trauma assessment. Major issues encountered when interpreting skeletal trauma include the influence of the deposition environment, endogenous and exogenous taphonomic processes, postmortem disturbance from scavengers, or relocation to secondary deposition sites [17–22]. All of these variables introduce the potential for trauma alterations that must be accounted for during skeletal analysis [17–22]. Although the variables influencing trauma interpretation differ from case to case, and even across elements of the same skeleton, the laws of bone biomechanics that guide these analyses stay constant [6,13–16]. The main consideration is that wet or living bone will respond to slow loading force (e.g., blunt force) by first absorbing the force through plastic deformation until the force overloads the bone causing it to fail (e.g., break) [6,11–14]. Plastic deformation is expressed in the bone as crushing of the cortical layer into the internal cancellous region, as the bone slowly absorbs force without exceeding its yield strength. Per contra, when a bone is exposed to rapid force, such as is seen with gunshot trauma, it will react as a more brittle material and fracture with little to no associated plastic deformation [6–8,13–16]. This brittle reaction is also characteristic of postmortem dry bone breakage [17–26]. Both plastic deformation and complete bone failure leave distinct signatures on the bone when observed both macro- and microscopically [6–8,17–26]. There is general agreement that if the biomechanics of bone's reaction to force remain as a constant variable, then interpreting the timing of traumatic injury should be possible despite postmortem taphonomic events and alterations [6–8,17–23].

Although the structural reactions of traumatized bone are well understood, post-depositional events can complicate interpretation. Taphonomic processes can introduce secondary fractures, alter fracture margins, or conceal impact sites [6,24–26]. Due to this, trauma signatures are addressed in variable depositional environments [1–11,15–28] to identify criteria that can be informative for trauma identification in specific contexts. However, few studies have addressed the influence of thermal alteration on blunt force trauma, specifically on the cranium. In forensic contexts, it is not uncommon for decedents to be disposed of by means of fire, as perpetrators of a crime often correlate the idea of a quick coverup with burning the body until only ash remains [20,21,24–28]. However, bodies exposed to fire burn slowly and are often recovered with intact skeletal or fleshed elements remaining [21,24–28]. When bodies are exposed to fire for a significant time, heat will alter the bone by degrading its organic components, leaving only the mineral structure [6,23–28]. The organic components, which are quickly dehydrated and destroyed from thermal modification, are what allow plastic deformation in living bone [6,12–15]. Therefore, thermal fractures express features similar to bone impacted by rapid force [6,17–22,25–28]. Further investigation is needed to understand the modifications caused by thermal exposure to perimortem trauma, as conclusions from the existing literature are unclear, inconsistent, and without validation [1,3–7]. Research derived for applications in forensic contexts is unique in its necessity for the method to pass the rigors required in legal proceedings [29–32]. All methods applied in forensic testing must be guided by strict sets of procedures and criteria. Thus, this research aims to identify characteristics that are indicative of fracture origin in thermally altered remains. Specifically, this paper highlights the characteristics of thermal and mechanically derived fractures of the cranium using *Sus scrofa* analogs, this being one of the most commonly traumatized regions in forensic contexts.

2. Materials and Methods

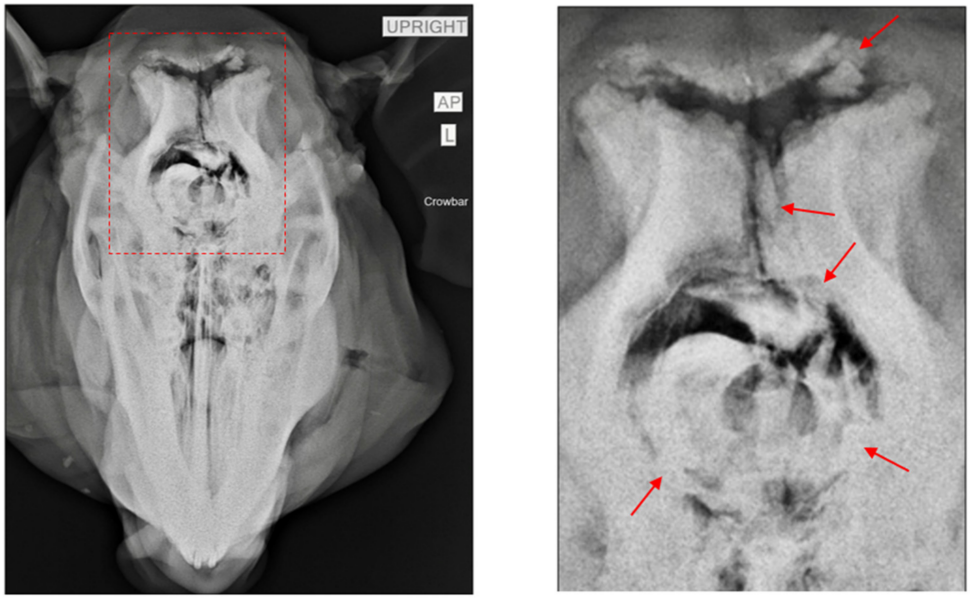
This study used 11 adult pigs as proxies for human remains, due to an established similarity in tissue thickness and structure between humans and pigs [2,3,11]. The pigs were procured from a local pork center and were humanely euthanized the morning of the

experiment with a blank bullet (following NC laws for humane slaughter) and shipped via cooling container to the pick-up site, where they were then transferred to the forensic laboratory. The samples consisted of skulls containing the complete cranium and mandible, and all were disarticulated from the axial skeleton prior to experimentation. No soft tissue was removed before experimentation, as removing tissue before blunt force trauma would not be consistent with an actual forensic event [24]. Of the 11 samples, 3 of the specimens were used as non-traumatized controls but were still subjected to burning. The remaining eight samples were divided into two groups, each containing four specimens (Table 1). One group was manually struck with a rounded crowbar and the other a flat-faced hammer. Two types of tools were used due to the different surface areas. Two samples of each group were traumatized with the head lying supine and the other two with the head positioned in the horizontal plane. This positioning allowed the recreation of forensic scenarios where a decedent is struck standing up (horizontal) or fallen (supine) with a buttressed surface creating secondary fractures opposite the initial impacts. Each specimen was struck on the frontal, zygomatic, parietal, and nasal bones. Samples were traumatized several times until fractures could be manually felt and then radiographed to document perimortem fracture patterns (Figure 1). Following radiographic documentation, the samples were taken to the burn site.

The burn site was located on the North Carolina State University dairy farm and the fire was constructed within a livestock feeding trough surrounded by cinderblock walls (Figure 2). An open-air, outdoor fire was implemented for this research, as this type of deposition is commonly encountered with forensic burning scenarios [1,5,24–26]. Materials involved in the creation and maintenance of the fire included wood logs and coals from previous fires. No accelerants were used in the process. Each sample was positioned with the head in the horizontal plane and maintained this position for the duration of the burn cycle (Figure 3). Specimens were placed directly on top of the logs in two rows, and documentation of the progressive thermal destruction was noted via photographs during the experiment. Total burn time was 1 h and 40 min, and the samples were burned until the calcined bone was seen across the samples. Once the degree of burning was sufficient to produce largely calcinated bone, the logs were removed from the fire to slowly decrease the temperature until the samples only remained on ash. The samples were left within the fire pit overnight to allow the specimens to completely cool before removal. The following morning the skulls and associated fragments were collected by hand from the pit, placed within individual containers, and returned to the laboratory for analysis.

Table 1. Sample distribution. Tool column indicates classification of instrument used during manual trauma. Position lists skull pose during trauma. Identifier # lists the classification system assigned to samples throughout experimentation.

Group	Tool	Position	Identifier #
1	Control-NA	NA	C1
	Control-NA	NA	C2
	Control-NA	NA	C3
2	Crowbar	Supine	CBS1
	Crowbar	Supine	CBS2
	Crowbar	Horizontal	CBH1
	Crowbar	Horizontal	CBH2
3	Hammer	Supine	HS1
	Hammer	Supine	HS2
	Hammer	Horizontal	HH1
	Hammer	Horizontal	HH2



(**a**) (**b**)

Figure 1. Radiographic image of sample CBH1. (**a**) Pre-burn radiograph of a specimen after manual trauma with a crowbar. Red dotted lines denote the area containing blunt force trauma. (**b**) Close-up image of the area within the red rectangle. Red arrows point to areas of incomplete fractures as a result of blunt force trauma, with associated fragments still attached.

Figure 2. Burn site.

Figure 3. Sample placement.

Laboratory analysis began with preprocessing photographs of each skull to document differential soft tissue destruction and note any thermal signatures before soft tissue removal. Following photography, any loosely adhering soft tissue was removed with a fine, soft-bristled brush. The skulls were reconstructed by refitting fragments using an adhesive. Control samples were analyzed first so that features of thermal fractures could be noted and established before comparison with the traumatized specimens. Location of fracture origin and termination, fracture type, skeletal color changes, and areas of soft tissue survival were recorded. Traumatized samples were reconstructed in the same manner as the controls and their pre-burning radiographs were compared post-burning (Figures 4 and 5).

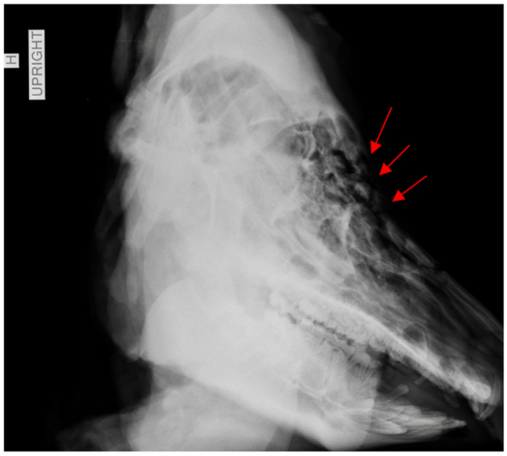


Figure 4. Radiographic image of specimen HH1. Red arrows point to areas of inwardly crushed bone, with displaced fragments.

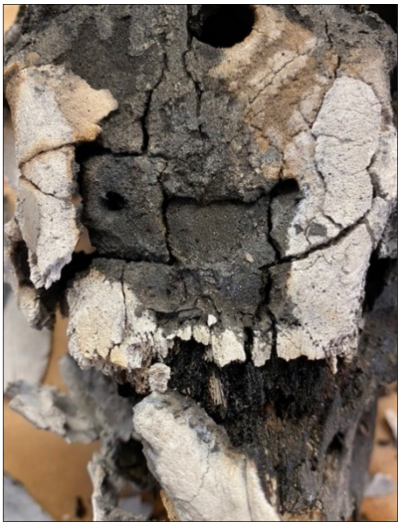


Figure 5. Post-processing reconstruction showing fragment association retaining depressed impact area.

3. Results

All three controls showed similar thermal alterations, which consisted largely of longitudinal fractures. These longitudinal fractures were inter-connected by transverse fractures or terminated transversely into an adjacent suture. In essence, thermal fractures appeared as patterns of long, rectangular fractures all over the cranium. Further, thermal fractures were found to be associated with cranial foramina and sutures. Thermal fracture propagation consistently originated from cranial sutures or foramina and terminated into longitudinal fractures or nearby sutures Figures 6 and 7. This finding was also highlighted in the study of Macoveciuc et al. (2017), who noted that due to the lack of accessory (traumatic) fractures in controls, heat accumulation caused fractures to originate from areas of the bone that could more easily vent, in this case being foramina and sutures. Thermal degradation was further characterized by cortical flaking and patina, a result of the rapid loss of organic components in the bone, and curved transverse fractures due to tissue regression [6,7,15,17–21].

Traumatized samples featured distinct characteristics observed only in the specimens that underwent mechanical force, which included the post-burning retention of plastic deformation, and impact areas that featured comminuted fracturing (Figure 8). Only green (e.g., wet, living) bone can respond to force as plastic deformation, as the impacted surface absorbs compressive force causing the opposite (internal) surface to tear from tension [6,12–14]. In all of the samples, blunt force fractures retained the pre-burning depressed areas and, in some samples, fragments were still connected to the associated fragment through an incomplete fracture. Due to the loss of the more pliable, organic components which allow for plasticity in bone, burning bone responds as a brittle material incapable of plastic deformation [6–9]. Only traumatized regions of the crania exhibited features of depressed fractures Figure 9. When considering fracture type, blunt force impacts were almost exclusively associated with comminuted fractures, a feature absent in controls or untraumatized regions. This difference in fracture type, being primarily longitudinal or comminuted, allowed for easier identification of suspect areas of trauma. The summary of fracture type and occurrence is presented in Table 2.

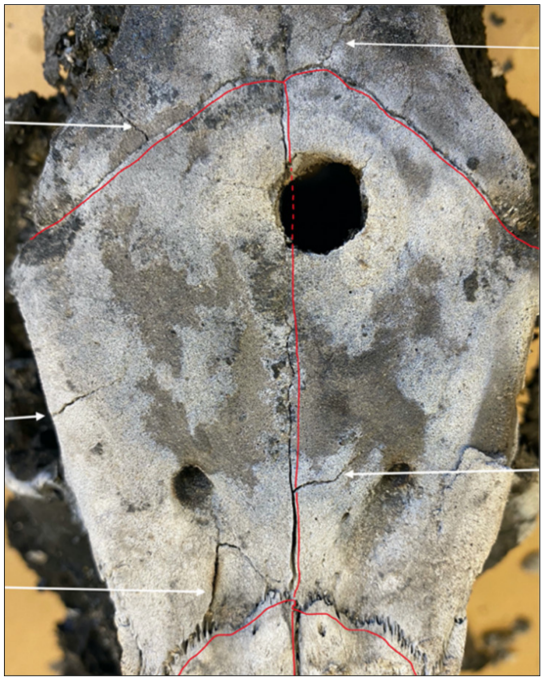


Figure 6. Red lines denote cranial sutures. White arrows point out thermal fractures originating and terminating within other sutures or foramina.

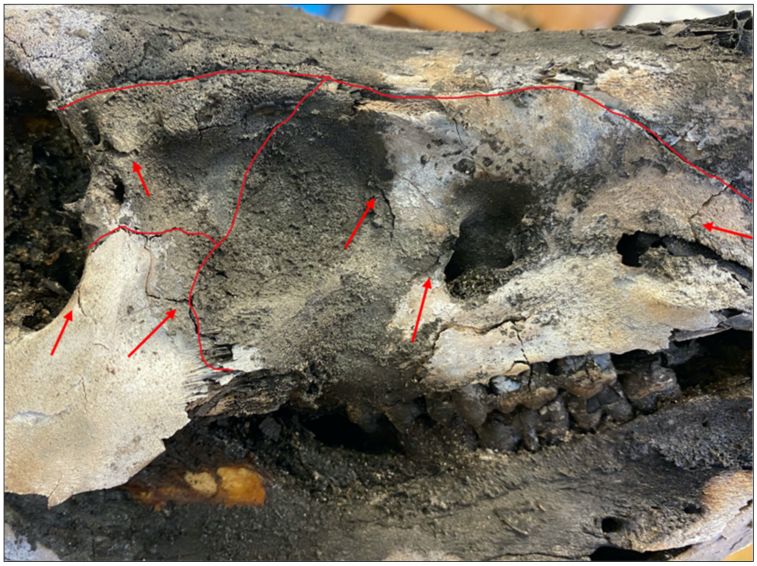
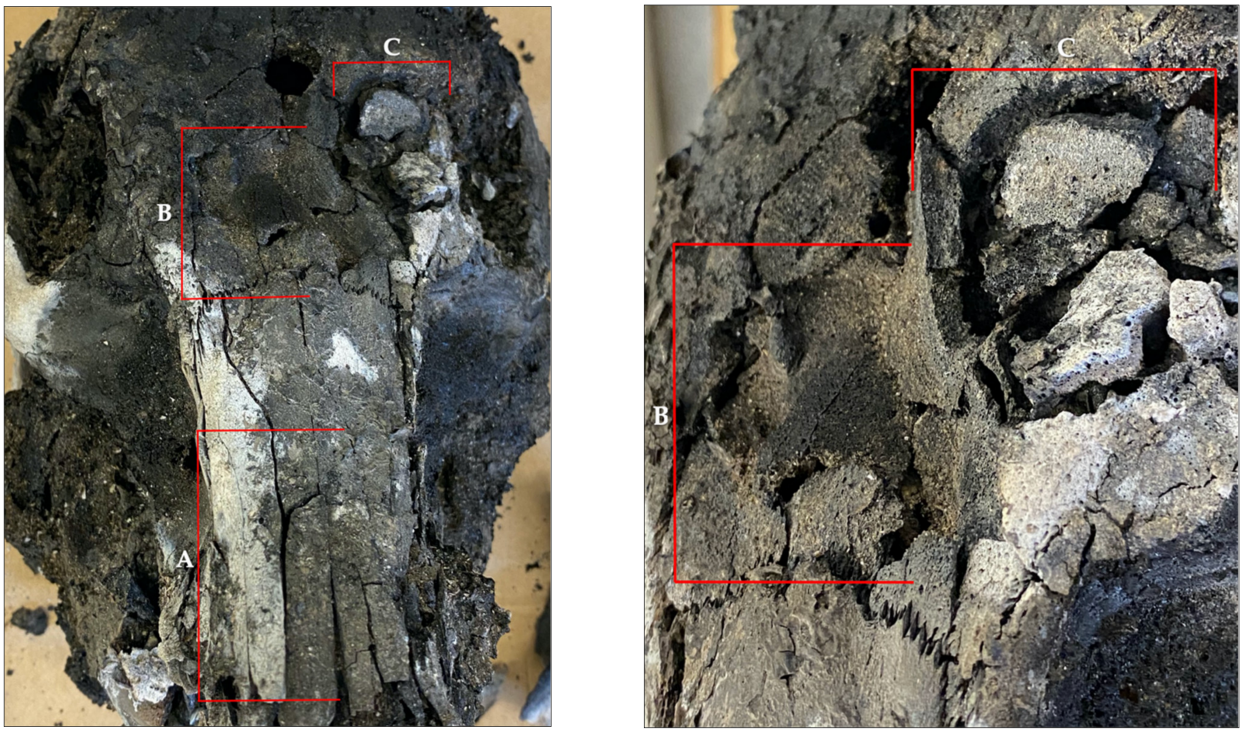
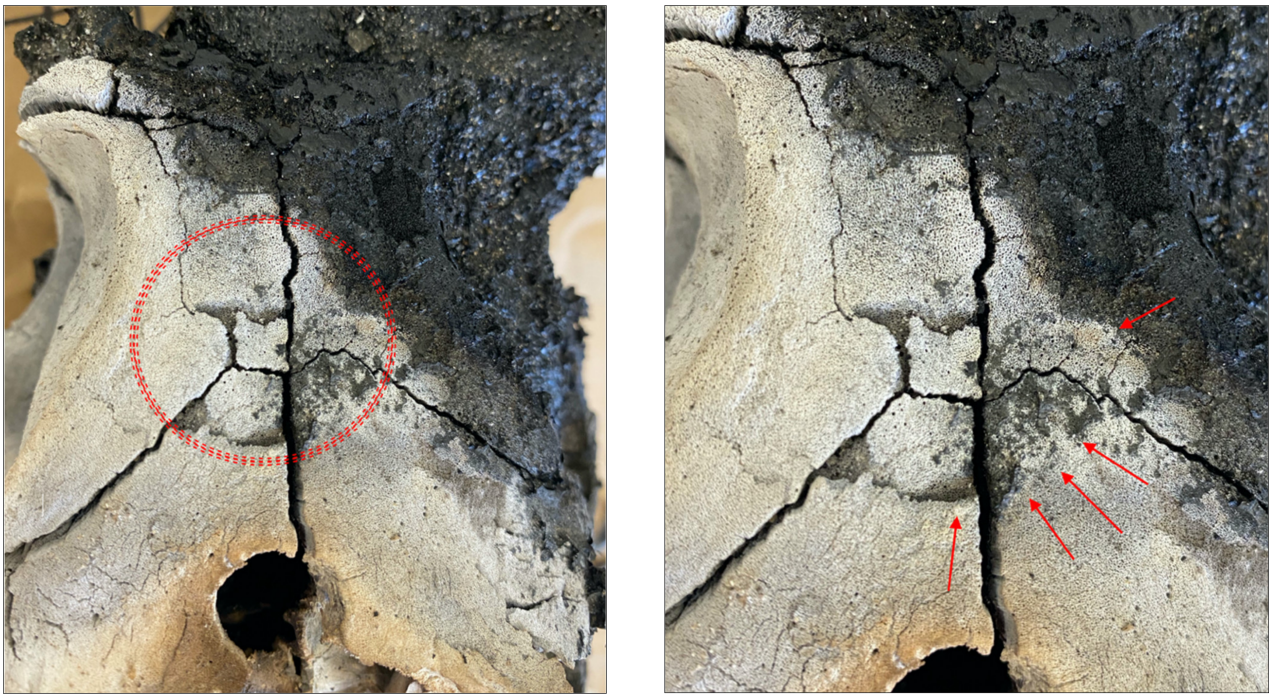


Figure 7. Red lines denote cranial sutures. Red arrows point out thermal fractures originating and terminating within other sutures or foramina.



(**a**) (**b**)

Figure 8. Sample CBH2 (**a**) **A** longitudinal thermal fractures interconnected with transverse fractures, **B** blunt force trauma retained showing depressed region with associated fragments, **C** comminuted fracture at impact site; (**b**) closer image of impact sites **B** and **C**.



(**a**) (**b**)

Figure 9. Hammer sample (**a**) region of impact exhibiting retained plastic deformation; (**b**) closer image of depression.

Table 2. Summary of fracture pattern observations. √ indicates presence of feature, X indicates absence of feature.

Sample	Defect Observed						
	Longitudinal	**Transverse**	**Comminuted**	**Curved Transverse**	**Depressed**	**Patina**	**Delamination**
C1	√	√	X	√	X	√	√
C2	√	√	X	√	X	√	√
C3	√	√	X	√	X	√	√
CBS1	√	√	√	√	√	√	√
CBS2	√	√	√	√	√	√	√
CBH1	X	X	√	√	√	√	√
CBH2	√	√	√	√	√	√	√
HS1	√	√	√	√	√	√	√
HS2	√	√	√	√	√	√	√
HH1	√	√	√	√	√	√	√
HH2	√	√	√	√	√	√	√

4. Discussion

The results of this pilot study demonstrated that consistent patterns of thermal alteration were noted that allowed for the differentiation of perimortem trauma after burning. Further, we did not find that thermal alterations obscured the blunt force trauma in any of the samples. Since this study incorporated the analyses of criteria noted to be of diagnostic value in similar research [1,3,6,7], the results of these analyses are discussed further.

4.1. Skeletal Biomechanics: Fracture Type & Morphology

Post-burning analyses found that the structural reactions between wet and brittle/dry bone were maintained, as fracture type and morphology reflected the material state of the bone when fractured. Plastic deformation was identified as areas of inwardly crushed bone with associated fragments still partially or completely attached. Although thermal exposure altered impact areas, exhibited as patina and flaking on fractured surfaces, the areas of impact retained the general morphology of the impact (as noted on pre-burn radiographs). In cases where no plastic deformation was retained, impact areas could be identified through the reassociation of fragments using an adhesive. Reassociated fragments displayed impact areas of clustered comminuted fractures (a feature absent in the thermally altered controls). Thermally altered controls consistently exhibited longitudinal, transverse, combination longitudinal-transverse, patina, and curved transverse (i.e., from tissue regression), but did not display any areas of comminuted patterns or depressions due to plasticity. Blunt force samples also presented these thermal alterations. However, regions of trauma were easily identified as variations from these thermal characteristics. In highly fragmented specimens, where the structural integrity of the cranium is lost and plastic deformation is absent, we find that identifying fracture patterns after fragment reassociation most consistently indicated the presence of trauma. Diagnostic importance has been given to fracture type and morphology in previous studies, and this study supports that these variables are indicative of fracture cause [1,3,6,7].

4.2. Fracture Origin and Termination

By first assessing the controls, location patterns of thermal fracture origin and termination were established. Thermal fractures appeared to originate and terminate in areas of the skull where thermal venting was present. That is, thermal fractures could be traced to cranial foramina or sutures and terminated within adjacent foramina and sutures. This finding is consistent with other studies that conclude that the pressure of high temperatures

within the cranium is released through natural vents, or openings, within the skull [1,6,7]. The pressure and heat released from this venting cause associated fractures to appear from these openings. The samples subjected to blunt force trauma showed the same fracture location patterns. However, locations of trauma deviated from this pattern as a cluster of comminuted fractures with no clear association to suture or foramina origins. Although fracture origin is variable when caused by traumatic force, thermal fractures are consistently associated with natural areas of thermal venting.

4.3. Skeletal Color Change

Previous suggestions [6,7] regarding color change as an indicator to differentiate fracture origin (e.g., due to thermal venting or burn progression) were not found to be useful in diagnosis for this experiment. Color changes appeared inconsistent in pattern or progression, exhibited as sporadic calcined islands surrounded by charred rings. Although potentially helpful for charting thermal degradation for remains consisting of a more complete body, these variables were not found to be of diagnostic value in this study. However, our study supports previous studies that fractures propagating into green or wet bone are associated with blunt force and perimortem trauma [1,6,7].

4.4. Tissue Thickness and Soft Tissue Survival

After evaluating body positioning and tissue thickness, it was observed that tissue regression and subsequent first areas of bone to burn reflected the general thickness of the tissue covering the bone, regardless of the position of the crania. The first areas to burn followed a pattern from the facial and snout regions, these being the least protected by muscle or tissue, with the thick tissues of the mandible burning last. The variability of tissue thickness across each skull created highly differential degrees of burning. Overall, facial regions were nearly calcinated while the mandible retained green bone under the lower facial muscles. Each skull consistently exhibited this pattern of tissue regression. After burning concluded, the only surviving tissues were those of the posterior portion of the mandible. Further, it appeared that fat acted as an accelerator of thermal destruction, while muscle acted as a protector. It was noted that regions of the skulls that were highly cartilaginous or fatty, such as the ears, burned more quickly than regions of the skull in more direct contact with the fire or with more densely concentrated muscle.

5. Conclusions

After incorporating analyses deemed to be of diagnostic value or indicative of blunt force trauma after thermal exposure in other studies, we found that the most valuable variable to identify the cause of fracture (e.g., blunt force or thermal alteration) is the fracture pattern. The result of this study found that tissue thickness is more indicative of thermal progression than body positioning and warrants further study when evaluating the progression of thermal destruction across skeletal elements.

Author Contributions: Conceptualization, K.K.; methodology, K.K. and A.H.R.; formal analysis, K.K.; writing—original draft preparation, K.K.; writing—review and editing, A.H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study met the requirements of U.S. Code Title 7, Chapter 48, Code §1902a: Humane Methods of Livestock Slaughter. This study was compliant with the ethical standards of the State of North Carolina.

Informed Consent Statement: Not applicable.

Data Availability Statement: Images will be made available upon request.

Acknowledgments: The authors would like to thank Nikki Long, Abdullah Zaben, Ginny Roman, Jordan Dalton, and Johann Galilea for their assistance with sample preparation and data collection.

We would also like to thank Wes Watson for his assistance with site preparation and fire management throughout the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Macoveciuc, I.; Márquez-Grant, N.; Horsfall, I.; Zioupos, P. Sharp and blunt force trauma concealment by thermal alteration in homicides: An in-vitro experiment for methodology and protocol development in forensic anthropological analysis of burnt bones. *Forensic Sci. Int.* **2017**, *275*, 260–271. [CrossRef] [PubMed]
2. Vaughan, P.; Vogelsberg, C.; Vollner, J.; Fenton, T.; Haut, R. The Role of Interface Shape on the Impact Characteristics and Cranial Fracture Patterns Using the Immature Porcine Head Model. *J. Forensic Sci.* **2016**, *61*, 1190–1197. [CrossRef] [PubMed]
3. Pope, E.; Smith, O. Identification of Traumatic Injury in Burned Cranial Bone: An Experimental Approach. *J. Forensic Sci.* **2004**, *49*, 431–440. [CrossRef]
4. Atherton, D.; McGwin, G.; Davis, G. Comparison of the Distribution of Cranial Fractures by Mechanism of Formation. *Acad. Forensic Pathol.* **2015**, *5*, 338–343. [CrossRef]
5. De Gruchy, S.; Rogers, T.L. Identifying chop marks on cremated bone: A preliminary study. *J. Forensic Sci.* **2002**, *47*, 933–936. [CrossRef] [PubMed]
6. Symes, S.A.; Rainwater, C.W.; Chapman, E.N.; Gipson, D.R.; Piper, A.L. Patterned thermal destruction of human remains in a forensic setting. In *The Analysis of Burned Human Remains*; Schmidt, C.W., Symes, S.A., Eds.; Academic Press: London, UK, 2008; pp. 15–54.
7. Herrmann, N.P.; Bennett, J.L. The differentiation of traumatic and heat-related fractures in burned bone. *J. Forensic Sci.* **1999**, *44*, 461–469. [CrossRef]
8. Delannoy, Y.; Becart, A.; Colard, T.; Delille, R.; Tournel, G.; Hedouin, V.; Gosset, D. Skull wounds linked with blunt trauma (hammer example). A report of two depressed skull fractures–Elements of biomechanical explanation. *Leg. Med.* **2012**, *14*, 258–262. [CrossRef]
9. Shipman, P.; Foster, G.; Schoeninger, M. Burnt bones and teeth: An experimental study of color, morphology, crystal structure and shrinkage. *J. Archaeol. Sci.* **1984**, *11*, 307–325. [CrossRef]
10. Richter, D.; Hahn, M.; Ostermann, P.; Ekkernkamp, A.; Muhr, G. Vertical deceleration injuries: A comparative study of the injury patterns of 101 patients after accidental and intentional high falls. *Injury* **1996**, *27*, 655–659. [CrossRef]
11. Bradley, A.; Swain, M.; Waddell, J.N.; Das, R.; Athens, J.; Kieser, J. A comparison between rib fracture patterns in peri- and post-mortem compressive injury in a piglet model. *J. Mech. Behav. Biomed. Mater.* **2014**, *33*, 67–75. [CrossRef]
12. Sorg, M. Differentiating trauma from taphonomic alterations. *Forensic Sci. Int.* **2019**, *302*, 109893. [CrossRef]
13. Gläser, N.; Kneubuehl, B.; Zuber, S.; Axmann, S.; Ketterer, T.; Thali, M.; Bolliger, S. Biomechanical Examination of Blunt Trauma due to Baseball Bat Blows to the Head. *J. Forensic Biomech.* **2011**, *2*, F100601. [CrossRef]
14. Wescott, D. Postmortem change in bone biomechanical properties: Loss of plasticity. *Forensic Sci. Int.* **2019**, *300*, 164–169. [CrossRef] [PubMed]
15. Snoeck, C.; Lee-Thorp, J.; Schulting, R. From bone to ash: Compositional and structural changes in burned modern and archaeological bone. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 55–68. [CrossRef]
16. Motherway, J.; Doorly, M.; Curtis, M.; Gilchrist, M. Head impact biomechanics simulations: A forensic tool for reconstructing head injury? *Leg. Med.* **2009**, *11*, S220–S222. [CrossRef]
17. Thompson, T. Recent advances in the study of burned bone and their implications for forensic anthropology. *Forensic Sci. Int.* **2004**, *146*, S203–S205. [CrossRef]
18. Thompson, T.J.U. Heat-induced dimensional changes in bone and their consequences for forensic anthropology. *J. Forensic Sci.* **2005**, *50*, 1008–1015. [CrossRef]
19. Galtés, I.; Scheirs, S. Differentiation between perimortem trauma and heat-induced damage: The use of perimortem traits on burnt long bones. *Forensic Sci. Med. Pathol.* **2019**, *15*, 453–457. [CrossRef] [PubMed]
20. Efford, M. The Implications of Thermogenic Modification for Anthropological Recovery of Burned Bone. *Arbutus Rev.* **2016**, *7*, 20. [CrossRef]
21. Viorel, P.; Mariana, R. Problems of forensic anthropological identification of carbonized human remains. *Rom. J. Leg. Med.* **2007**, *15*, 39–44. [CrossRef]
22. Bohnert, M.; Rost, T.; Faller-Marquardt, M.; Ropohl, D.; Pollak, S. Fractures of the base of the skull in charred bodies—post-mortem heat injuries or signs of mechanical traumatisation? *Forensic Sci. Int.* **1997**, *87*, 55–62. [CrossRef]
23. Spennemann, D.H.R.; Colley, S.M. Fire in a pit: The effects of burning of faunal remains. *Archaeozoologia* **1989**, *3*, 51–64.
24. Fanton, L.; Jdeed, K.; Tilhet-Coartet, S.; Malicier, D. Criminal burning. *Forensic Sci. Int.* **2006**, *158*, 87–93. [CrossRef] [PubMed]
25. Bradtmiller, B.; Buikstra, J. Effects of Burning on Human Bone Microstructure: A Preliminary Study. *J. Forensic Sci.* **1984**, *29*, 11701J. [CrossRef]
26. Egeland, C.; Pickering, T. Cruel traces: Bone surface modifications and their relevance to forensic science. *Wires Forensic Sci.* **2021**, *3*, e1400. [CrossRef]

27. Tümer, A.; Akçan, R.; Karacaoğlu, E.; Balseven-Odabaşı, A.; Keten, A.; Kanburoğlu, Ç.; Unal, M.; Dinç, A.H. Postmortem burning of the corpses following homicide. *J. Forensic Leg. Med.* **2012**, *19*, 223–228. [CrossRef] [PubMed]
28. Ubelaker, D. The forensic evaluation of burned skeletal remains: A synthesis. *Forensic Sci. Int.* **2009**, *183*, 1–5. [CrossRef]
29. Christensen, A.; Crowder, C. Evidentiary Standards for Forensic Anthropology. *J. Forensic Sci.* **2009**, *54*, 1211–1216. [CrossRef]
30. Lesciotto, K. The Impact of Daubert on the Admissibility of Forensic Anthropology Expert Testimony. *J. Forensic Sci.* **2015**, *60*, 549–555. [CrossRef]
31. National Research, Council; National Academy of Sciences. *Strengthening Forensic Science in the United States: A Path Forward [Hereinafter NAS Forensic Science Report]*; The National Academies Press: Washington, DC, USA, 2009.
32. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579.1993. Available online: https://supreme.justia.com/cases/federal/us/509/579/ (accessed on 20 December 2021).

biology

MDPI

Article

The Effects of Cranial Orientation on Forensic Frontal Sinus Identification as Assessed by Outline Analyses

Lauren N. Butaric *, Allison Richman and Heather M. Garvin

Department of Anatomy, Des Moines University, 3200 Grand Avenue, Des Moines, IA 50312, USA; Allison.Richman@dmu.edu (A.R.); Heather.Garvin-Elling@dmu.edu (H.M.G.)
* Correspondence: Lauren.Butaric@dmu.edu

Simple Summary: Frontal sinus patterns are unique amongst individuals. When faced with an unknown decedent, investigators can compare the frontal sinus pattern observed in postmortem radiographs to antemortem radiographs of the suspected individual to make a positive identification. Ideally, the antemortem and postmortem radiographs are oriented in the same exact position, but this can be challenging. This study investigates how slight variations in radiographic orientation affect sinus outlines and potentially impact identification. Frontal sinus models were created from CT scans (21 individuals) and digitally oriented across three clinically relevant views. From each standard orientation (looking straight ahead), eight 5° deviations were obtained in horizontal (left/right), vertical (up/down), and diagonal (e.g., left-up vs. right-down) directions. Within and between individual differences in sinus size and outline shape were assessed. Sinus breadth remained relatively stable across deviations, while sinus height was affected by small vertical deviations. Although radiographic vertical deviations resulted in statistical differences, impacts on outline matches were minimal. However, practitioners need to take particular care in matching radiographic orientation for smaller and/or discontinuous (right and left sides separated) sinuses, which are more likely to lose part of the sinus in more inferiorly oriented views and, thus, could affect various methods of sinus identification.

Abstract: The utility of frontal sinuses for personal identification is widely recognized, but potential factors affecting its reliability remain uncertain. Deviations in cranial position between antemortem and postmortem radiographs may affect sinus appearance. This study investigates how slight deviations in orientations affect sinus size and outline shape and potentially impact identification. Frontal sinus models were created from CT scans of 21 individuals and digitally oriented to represent three clinically relevant radiographic views. From each standard view, model orientations were deviated at 5° intervals in horizontal, vertical, and diagonal (e.g., left-up) directions (27 orientations per individual). For each orientation, sinus dimensions were obtained, and outline shape was assessed by elliptical Fourier analyses and principal component (PC) analyses. Wilcoxon sign rank tests indicated that sinus breadth remained relatively stable ($p > 0.05$), while sinus height was significantly affected with vertical deviations ($p < 0.006$). Mann–Whitney U tests on Euclidean distances from the PC scores indicated consistently lower intra- versus inter-individual distances ($p < 0.05$). Two of the three orientations maintained perfect (100%) outline identification matches, while the third had a 98% match rate. Smaller and/or discontinuous sinuses were most problematic, and although match rates are high, practitioners should be aware of possible alterations in sinus variables when conducting frontal sinus identifications.

Keywords: human identification; frontal sinus shape; outline analysis; elliptical Fourier analysis; computed tomography; radiology

Citation: Butaric, L.N.; Richman, A.; Garvin, H.M. The Effects of Cranial Orientation on Forensic Frontal Sinus Identification as Assessed by Outline Analyses. *Biology* **2022**, *11*, 62. https://doi.org/10.3390/biology11010062

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 22 November 2021
Accepted: 29 December 2021
Published: 2 January 2022

1. Introduction

The potential of radiographic comparisons and forensic identification based on frontal sinus morphology, in particular, has been recognized since the 1920s [1–18]. Frontal sinus morphology is highly individualized with differences noted even between monozygotic twins, which makes them an ideal candidate for identification [2–4,19]. If antemortem radiographic images containing the frontal sinus are available for a suspected decedent, a comparison with postmortem radiographs can provide a fast and inexpensive method of identification, similar to radiographic dental comparisons. Comparisons can also frequently be made on fragmented or even burned remains. Specific methods for radiographic frontal sinus identification range from qualitative visual assessments [1,3,19,20] to the use of metrics and/or coded traits [7,8,17,21,22], and outline analyses [15,16,23]—all of which rely at least partly on sinus shape as defined by the presence of right/left sinus lobes, as well as the individual scallops and arcades that give the sinus their distinctive outlines.

All frontal sinus radiographic methods also require a postmortem radiograph taken in the same orientation as the antemortem record. Antemortem radiographs reflect standard clinical views typically used in sinus or head-and-neck imaging, with three common orientations being Caldwell's, posterior–anterior (PA)-frontal, and Water's view [24]. Forensic anthropologists are more familiar with the Frankfort Horizontal orientation of the cranium and may be inclined to take postmortem radiographs in this anthropological orientation; further, several studies on frontal sinus variation utilize this orientation [25,26]. These orientations vary in the positioning of the head/cranium relative to the film and the trajectory of the X-rays. As such, the appearance of the frontal sinus on the two-dimensional radiographs may be altered based on the radiographic orientation chosen.

Given the potential effects of orientation on radiographic representation, practitioners acquiring postmortem radiographs should aim to position the cranium in a similar orientation as the antemortem view. Still, obtaining a perfect alignment, however, can be challenging. Owing to this, several previous studies have investigated how slight variations in skull orientation affect the radiographic presentations of frontal sinus morphology and individual identification methods [25–29]. Overall, studies suggest that even 5° degrees of varying orientation may affect sinus morphology. However, these previous studies are limited in scope. For example, when testing their outline-based method, Christensen [28] was only able to effectively measure two crania in a single clinical view. Silva et al. [25] investigated how varying 10° vertical orientations affects frontal sinus breadth, but only incorporated a single individual and limited analyses to PA-frontal view. It is possible that the degree of error introduced is dependent on sinus size and complexity. So, the degree of shape deviation obtained from these limited studies may depend on the frontal sinus morphology of the single individual used in the error analysis.

In each of the above studies, the physical placement and repositioning of the crania was done by the technicians and could add human error/bias given the challenges of obtaining perfect alignments. In fact, Hashim et al. [29] found that repositioning of crania on radiographic tables, even after only a short time has passed between repositioning (less than one minute) results in significantly different sinus presentations. To account for this, Riepert et al. [27] utilized a specialized program that simulates radiographs from CT-derived data. These authors digitally re-oriented crania in 4° and 8° variations. Ultimately, they found that frontal sinus breadth and height presented with high degree of variability across the orientations, but the inter-individual uniqueness of the sinus was such that these variations did not result in misidentifications. More recently, Nikolova et al. [26] utilized an industrial μCT scanner to obtain radiographic images, which allowed re-orientation of crania using more precise means via a computer-automated tilting gantry. They compared linear measures of the frontal sinus across 10 orientations at 5° intervals, starting from the Frankfort Horizontal plane at −20° to a view at 45° with the midpoint being 0°. Overall, they found significant differences in height and breadth measures at 5° vertical variations from the 0° midpoint. While the use of an industrial scanner allowed hands-free vertical

tilting of the cranium, they were unable to incorporate lateral movements. Further, they did not test for implications of these findings to forensic identification methods.

There is a need to more thoroughly understand how minor deviations in radiographic orientation may affect forensic frontal sinus identifications, given the severe consequences of a mistaken identification or erroneous exclusion of identity. This study aims to assess how 5° vertical, horizontal, and diagonal (i.e., combined 5° vertical and horizontal) deviations in orientation from standard views affect frontal sinus shape as captured by outlines using a sample of computed tomography (CT) scans from 21 individuals. The increased sample size, inclusion of horizontal, vertical, and diagonal deviations, and testing of three standard radiographic orientations builds upon the previous literature, providing a more comprehensive study and the ability to directly compare results from the same sample across the orientations and deviations. This study also tests how these varying orientations may directly affect forensic identification, with a focus on the frontal sinus outline method devised by Christensen [15,16,28]. This method was chosen as it is one of the most cited methods for forensic sinus identification [30] and provides a means of capturing overall frontal sinus shape, which can then be quantitatively analyzed. The results of this study will help guide best practices in forensic frontal sinus identifications.

2. Materials and Methods

2.1. Materials

This study utilizes computed tomographic (CT) scans originating from the Robert J. Terry Anatomical Collection, National Museum of Natural History, Smithsonian Institution (Washington, DC, USA) [31]. The current sample included 21 adult crania (aged 20 to 95, average age = 51.667), with 13 African American (7 females, 6 males) and 8 European American (6 males, 2 females) individuals. The 21 crania were selected from a larger sample of CT scans publicly available from Lynn Copes' website [32,33]. Only individuals possessing frontal sinuses above the supraorbital line, with no obvious signs of pathologies affecting the frontal sinus were utilized. Additional sample information and scanning protocols are provided by Copes [32,33]. Although the sample does not encompass broad ranges of temporal, population, or specific age variations, the crania included displayed a wide range of sinus size and morphology. Thus, this methodological study is able to assess effects of deviations on a broad range of individual sinus morphologies, as appropriate for our question. Although some studies have documented patterns in frontal sinus morphology between population, sex, or age groups or with variables such as body size and craniofacial morphology, these relationships are relatively weak [34–36], and the underlying factors contributing to such a high degree of frontal sinus variation, whether within or across groups, are unknown. These variations are described as differences in sinus appearance, size, and shape— all variables included in the present study. Thus, although this study does not include a highly diverse sample, the results of the analyses should be applicable across groups.

For this study, the CT scans were imported into the program Amira5.6 [37], where semi-automatic processes were used to segment the frontal sinus, effectively creating a virtual endocast, and model the cranium following a previous study [38]. Both objects were digitally rendered and saved as two stereolithographic (.stl) models. Although there were two models, they maintained the same coordinate space and could be manipulated (e.g., oriented) together as if they were a single object. Care was taken not to employ smoothing techniques or any processing techniques that would alter frontal sinus morphology.

2.2. Sinus Orientations

The aim of the study was to assess how minor deviations from common radiological views could affect the observed sinus morphology and how that may impact forensic identification methods. When deciding which radiological orientations to include in the study, common clinical and anthropological radiographic orientations (i.e., Frankfort Horizontal, PA-frontal, Caldwell's, Water's view) were considered. Specific definitions

of these clinical views can vary by source, with some definitions focusing on soft tissue structures (e.g., nose against film) or the resultant radiographs (e.g., alignment of the petrous portions within the orbits), instead of specific osteometric landmarks. Clinical orientations for head radiography depend on both the positioning of the head relative to the X-ray film, as well as the angle of the beam trajectory. Given that this study utilizes CT scans to reduce subjectivity in positioning of these minor deviations in orientation, we could not emulate changes in beam trajectory (i.e., the CT scans are most similar to an X-ray beam trajectory perpendicular to the film); thus, this limitation was also considered when choosing which radiographic views to test. Priority was also given to orientations used in past frontal sinus and orientation studies for comparative purposes.

Given the above considerations, three radiographic planes were chosen for evaluation: Frankfort Horizontal, Orbitomeatal, and Porion-Alveolar (see Figure 1). The Frankfort Horizontal Plane (FHZ) was chosen for several reasons: it is used in several clinical settings, particularly for occlusal and temporomandibular evaluations [39–42]; it is a common orientation in frontal sinus identification research [7,43–45]; it has been utilized in previous studies specifically investigating the role of varying orientations on frontal sinus morphology [25–27]; and it is most familiar to forensic anthropologists. For FHZ, crania are oriented such that the left-sided landmarks of *porion* (superior aspect of the external auditory meatus) and *orbitale* (most inferior margin of the orbit) are aligned in one axial plane, with left and right *porion* as level as possible (Figure 1, middle). For this study, FHZ was considered as the intermediate view, as the remaining views alter the crania inferiorly and superiorly relative to FHZ.

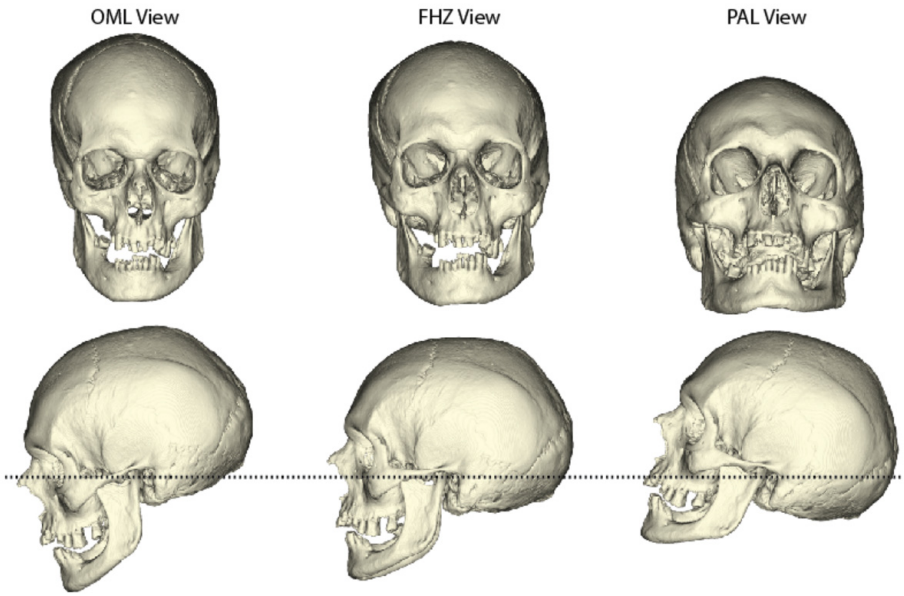


Figure 1. Standard cranial views utilized in this study from left to right: Orbitomeatal Line (OML), Frankfort Horizontal Plane (FHZ) and Porion-Alveolar Line (PAL). Dashed line represents the axial plane of orientation, see text for details.

The Orbitomeatal Line (OML) was chosen as it is commonly referenced in clinical radiographic views, such as PA-frontal views, with X-ray beam trajectories following this axis and perpendicular to the film. Thus, not only is this view easy to replicate with CT scans, but antemortem radiographs obtained for forensic identification may frequently be in this position due to their use for evaluating the midfacial regions. For this view, crania are oriented such that the center of the external auditory meatus and the middle of the orbital cavity are aligned in the same axial plane, with left and right sides as level as possible. Note, Cruz and Gasperini [46] found that the OML is approximately 15° from FHZ, with the cranium rotated more inferiorly in OML.

The third view, which we termed the Porion-Alveolar Line (PAL), was primarily chosen for comparative purposes, as it represents a superior rotation of the head and is utilized in several previous studies investigating the effect of orientation on frontal sinus morphology. Following Silva et al. [25] and Nikolova et al. [26], we obtained this view by rotating the cranium 20° superiorly from FHZ (Figure 1, right). It is important to note, however, that these previous studies misleading refer to this orientation as "Caldwell view," which does not match the clinical definition. Clinically, Caldwell's view is obtained by having the patient put their forehead and nose against the X-ray film and then orienting the X-ray beam at a 15–20° angle to the film [24]. In this orientation, the head position resembles the OML view, but the X-ray beam traverses from the occipital bone (near *lambda*) through the mid-orbit region. The superior rotation of the cranium 20° from FHZ (as done by Silva et al. [25] and Nikolova et al. [26]), rotates the head in the opposite direction as the clinical Caldwell head orientation and results in an X-ray beam trajectory passing through the *porion* and the maxillary alveolus, very different from that of actual Caldwell's view. Further, the end result of Caldwell's view should consist of the petrous pyramids located in the lower third portion of the orbits [24], which is not evident in the figures provided by Nikolova et al. [26]. While we find this view informative for reasons below, we do not use the "Caldwell" notation in this study. Instead, we refer to it as the Porion-Alveolar Line (PAL), as this more accurately reflects the beam trajectory through the cranium. Despite not being defined as a typical clinical view, we include PAL here for comparative purposes given that these previous studies used this orientation, found that it was the most stable in terms of frontal sinus morphology across varying vertical orientations, and provided the clearest view of the sinus [25,26]. Its incorporation also provides additional insight into how superior vertical inclination of the cranium may affect frontal sinus morphology. Given the relationship of FHZ and OML axes, the PAL orientation can be inferred to be approximately 30° superiorly rotated from the OML line. While Water's view, defined as 45° superior rotation from the OML view, is a clinical view that could have been investigated, it was not specifically tested in this study given that its extreme superior rotation has already been shown to be highly susceptible to vertical deviations [24,27,47]. Further, preliminary investigations in the current study indicated that Water's view would result in several instances where the frontal sinus would be eliminated completely from view. The PAL view provides a test of a less-extreme vertical orientation.

2.3. Frontal Sinus Outlines

The associated frontal sinus and cranial models for each individual were imported into the program 3DSlicer [48]. The models were oriented into these three main orientations (FHZ, OML, PAL) using cranial landmarks, and then eight additional views varying in 5° intervals from each of the main orientations, resulting in 27 radiographic views per individual. The Transformation module was used to digitally rotate the models to the defined degrees and alleviate human error in obtaining the 5° rotations. Within each view, the 5° deviations were defined from the standard orientation as horizontal deviations (5° left; 5° right), vertical deviations (5° up; 5° down), and diagonal deviations (5° left and up; 5° right and up; 5° left and down; 5° right and down). These are illustrated in Figure 2; note the cranial model is included for interpretative purposes only and was not included when obtaining sinus outlines as described below.

Following previous studies [15,16,23,28], the inferior border of the frontal sinus was demarcated at the level of the superior orbital margin in each view; the sinus remaining below the line was deleted from view using the Model Clipping tool in 3DSlicer. Once oriented correctly with the inferior border demarcated, the cranial model was hidden from the view, leaving the properly oriented and clipped sinus model above the superior orbital margin. The background was set as black, the sinus model set at white, and a 10 cm scalebar was added. A two-dimensional (2D) image was then captured of the sinus and scalebar using the Annotations Screen Capture module in 3DSlicer. This resulted in a total of 567 images for 21 individuals.

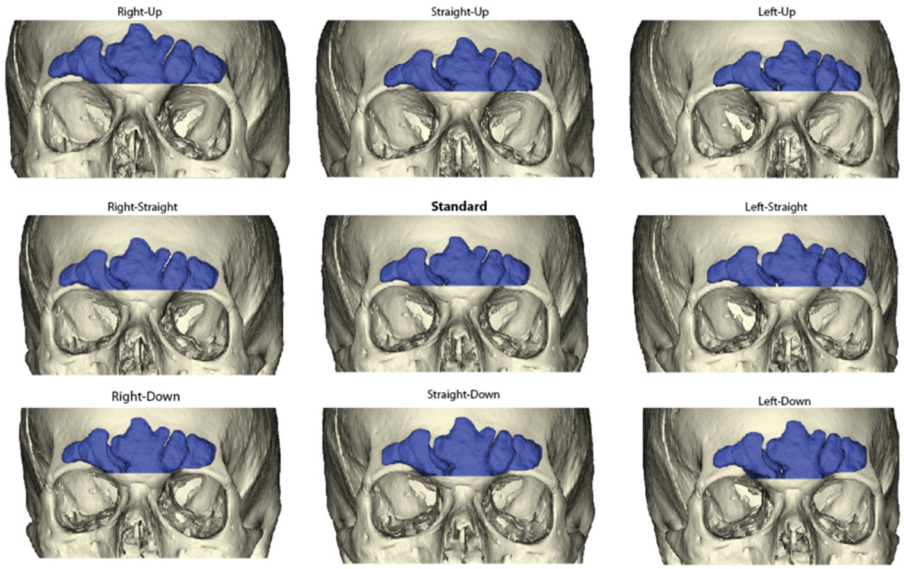


Figure 2. Graphical representation of the nine varying 5-degree orientations used in each view. Frankfurt Horizontal Plane pictured. Note actual 2D images used for analyses did not include the cranium.

The 2D images were then imported into ImageJ [49] where they underwent further processing. First, image sizes were increased to 3000 pixels. Given that outline analyses require a single continuous outline, if the right and left lobes of the frontal sinus were separated (i.e., discontinuous) they were connected by a white line (set as 2 pixels thick). Additionally, each image was scaled according to the 10cm scale bar obtained from 3DSlicer, and frontal sinus area, maximum breadth (taken parallel to the supraorbital line), and maximum height (taken perpendicular to the supraorbital line) were collected using the Measurement tools in ImageJ. Maximum breadths and heights were taken twice by the same observer and then averaged.

Following Christensen [15,16,28], frontal sinus outlines were based on the external contour of the sinus, with the supraorbital line demarcating the inferior boundary. Elliptical Fourier analysis (EFA) was conducted on the outlines to quantitatively capture outline shape as a series of harmonics. Unlike other morphometric methods, EFA does not require homologous landmarks (which the frontal sinus lacks); instead, the outline shape is captured by harmonics, and resultant elliptical Fourier coefficients can be used to assess sinus shape (see [50] for a general review of EFA in forensic anthropology). EFA analyses were conducted using the SHAPE software [51] where the frontal sinus outlines were automatically digitized based on the 2D images of the white sinus models against the black background. The outlines were converted into numeric codes, referred to as "chain codes," using the CHC module. Next, the CHC2NEF module was used to convert the codes into elliptical Fourier coefficients represented by 20 harmonics and normalized by the first harmonic. The resulting coefficients were then subjected to a principal component analysis (PCA) using the Princomp module, and the effective PCs (i.e., those with proportions larger than $1/n_{\text{coefficients}}$) were retained for all subsequent statistical analyses.

2.4. Statistical Analyses

Unless otherwise noted, all analyses were conducted in SPSS v28 [52], using a significance of 0.05. Initial exploration of the data indicated several measures violated the assumption of normality (i.e., Shapiro–Wilks p-values < 0.05). To be conservative, non-parametric statistics were utilized for all analyses. Spearman's Rho correlation analyses were conducted to gain initial insights into how the PCs varied. To test for significant differences in sinus morphology due to varying orientations within each of the three views, Wilcoxon sign ranked tests were conducted. Owing to the assumption that investigators

would place crania as close to the antemortem image as possible, and in order to reduce the number of pair-wise comparisons, we focused our analyses on the deviations from the standard orientation within each view, versus investigating orientations across the three major views. Specifically, the effects of the 5° variations on sinus variables (i.e., area, breadth, height, PCs) in the standard FHZ, OML, and PAL views were tested against their eight respective deviations in orientation. Due to multiple tests in this section, we applied a Bonferroni correction: significant differences were considered at the 0.006 alpha-level (0.05 divided by eight tests, per standard view). Descriptive statistics and plots were used to interpret differences and trends observed between the views and deviated orientations. Visualizations of the PCs and original sinus outlines were also utilized when interpreting how the orientations affected sinus shape.

The final portion of this study was to test the implication of varying orientation on frontal sinus identification. To accomplish this, multivariate Euclidean distances were calculated across the effective PCs between each varying orientation, both within individuals (intra-individual distances) and among different individuals (inter-individual distances). The assumption was that the intra-individual distances should be significantly less than the inter-individual distances. Three Euclidean distance matrices—one for each view—were created using the program PASSaGE2.0 [53]. Mann–Whitney U tests were conducted to statistically compare the pooled intra-individual distances (all orientations; n = 36 distances per individual, with 756 distances per view) to pooled inter-individual distances (all orientations; n = 1620 per individual, with a total of 34,020 distances per view). Using a one-tailed hypothesis, these analyses allowed us to directly test whether inter-individual distances were significantly greater than the intra-individual distances within each view. To test whether a specific view (FHZ, OML, or PAL) was less/more reliable than others, a Kruskal–Wallis analysis was conducted to test for significant differences in the intra-individual distances among the three views. If significant, follow-up Mann–Whitney U tests were conducted to directly test for specific differences among the three views. These analyses were conducted in SPSSv28 using a significance of 0.05, unless otherwise noted.

Finally, to assess whether the slight deviations from standard orientations could affect forensic frontal sinus matches as based on Christensen's outline method [15,54], we also took each outline (n = 189; 21 individuals and 9 outline views) and determined which outline they most closely matched to (i.e., least Euclidean distance). This was carried out within each of the standard views (i.e., all FHZ compared to all other FHZ outlines). If the deviations in orientations do not grossly affect positive identification, then the smallest distance (closest match) should be to an outline within the same individual and not an outline from a different individual.

3. Results

The PCA yielded eight effective PCs explaining a cumulative 92.57% of the variation. The results of Spearman's Rho correlation analyses between the PCs and sinus variables are presented in Table 1. In terms of the sinus dimensions, PC1 most closely approximates sinus height, as indicated by the higher correlation coefficients, compared to PC2 and PC3. More specifically, PC1 (35.77% of the variation) largely tracked height and breadth dynamics; individuals with negative PC1 scores expressed superior inferiorly flatter sinuses (i.e., relatively larger breadth than height) compared to positive PC1 scores. PC2 (24.20%) appears to capture sinus complexity; outlines with more negative PC scores have several large "loops" and "indentations", with some of the indentations approaching the supraorbital line, while those with more positive PC scores lack these indentations. PC3 (13.53%) tracks relative height changes in the outline, with negative PC3 scores representing sinuses with a distinctly higher midline (i.e., similar to a mountain peak) compared to lateral areas and positive PC3 scores representing sinuses with more equally distributed heights across the outline (i.e., similar to a plateau). The remaining PCs each explain less than 10% of the variation and were not univariately analyzed further.

Table 1. Results of Spearman's Rho correlation analyses (with correlation coefficients and *p*-values) between sinus variables and the principle components (PCs) representing >10% of variation.

	Area	**Breadth**	**Height**	**PC1**	**PC2**	**PC3**
Area	-	0.900	0.921	0.666	0.465	0.067
p-value	-	**<0.001 ***	**<0.001 ***	**<0.001 ***	**<0.001 ***	0.111
Breadth	0.900	-	0.873	0.615	0.295	−0.249
p-value	**<0.001 ***	-	**<0.001 ***	**<0.001 ***	**<0.001 ***	**<0.001 ***
Height	0.931	0.873	-	0.840	0.380	0.019
p-value	**<0.001 ***	**<0.001 ***	-	**<0.001***	**<0.001 ***	0.658
PC1	0.666	0.615	0.840	-	0.092	0.068
p-value	**<0.001 ***	**<0.001 ***	**<0.001 ***	-	**0.029 ***	0.104
PC2	0.465	0.295	0.380	0.092	-	0.109
p-value	**<0.001 ***	**<0.001 ***	**<0.001 ***	**0.029 ***	-	**0.009 ***
PC3	0.067	−0.249	0.019	0.0368	0.109	-
p-value	0.111	**<0.001 ***	0.658	0.104	**0.009 ***	-

* Bold text with asterisk indicates significance at the alpha level of 0.05.

Table 2 provides descriptive statistics for PCs 1–3 in each orientation, along with the Wilcoxon sign ranked tests between the standard (i.e., non-deviated) view and each 5° variation of that view. PC1 and PC3 present with more significant differences compared to PC2, which likely relates to their relationship with size-related (i.e., relative height and breadth) shape changes. All significant differences occur within the FHZ and OML views and all significant differences involve some deviation in a vertical component (i.e., left-straight and right-straight deviations did not result in any significant differences). As individuals are oriented further inferiorly, they tend to display significantly shorter heights relative to breadth (i.e., more negative-loading PC1 and PC3s). This can be seen in Figure 3, which shows PC1–PC3 values (with associated contours) against 5° vertical and horizontal deviations (diagonal views not pictured).

The interpretations of the PC results were confirmed by analyzing the sinus area, height, and breadth measurements. Table 3 provides descriptive statistics for these variables in each orientation, along with the Wilcoxon sign ranked tests between the standard and each varying 5° orientation within each view (e.g., FHZ standard vs. FHZ left up). OML and FHZ displayed significant differences in area and height for most views, while PAL only displayed significant differences in height. Again, deviations without any vertical component (i.e., right straight and left straight) did not result in significant differences. There were no significant differences in breadth among the 5° deviations and any of the three standard views. These results are best illustrated in Figure 4, which shows individual area (top graph), height (middle graph), and breadth (bottom graph) dimensions against 5° vertical and horizontal orientations (diagonal views not pictured). Note the relatively stable breadth dimensions, with one exception in the OML down view (see discussion).

Table 2. Median and inter-quartile range (IQR) statistics for PC1–PC3 across the three views and varying orientation. Wilcoxon sign rank tests (Z scores and significance) also provided for each orientation versus respective standard view.

Orientation	**PC1**		**PC2**		**PC3**	
	Median (IQR)	**Z**	**Median (IQR)**	**Z**	**Median (IQR)**	**Z**
OML Standard	−0.075 (0.187)	–	0.009 (0.109)	–	−0.044 (0.094)	–
Straight Down	−0.090 (0.184)	**−2.798 ***	0.005 (0.082)	−0.261	−0.042 (0.097)	−2.450
Straight Up	−0.055 (0.174)	**−3.424 ***	0.022 (0.109)	**−3.076 ***	−0.018 (0.069)	**−3.389 ***
Right Straight	−0.079 (0.191)	−1.477	−0.004 (0.128)	−2.138	−0.046 (0.090)	−1.373
Right Down	−0.058 (0.171)	**−2.798 ***	0.040 (0.144)	−1.651	−0.015 (0.065)	**−2.768 ***
Right Up	−0.065 (0.177)	**−3.076 ***	0.004 (0.071)	−0.052	−0.052 (0.076)	**−2.763 ***

Table 2. *Cont.*

Orientation	PC1		PC2		PC3	
	Median (IQR)	**Z**	**Median (IQR)**	**Z**	**Median (IQR)**	**Z**
Left Straight	−0.068 (0.185)	−1.408	0.005 (0.123)	−0.226	−0.030 (0.077)	−2.450
Left Down	−0.065 (0.187)	−2.555	0.017 (0.133)	−1.651	−0.009 (0.061)	**−3.250 ***
Left Up	−0.099 (0.163)	**−3.111 ***	0.008 (0.095)	−0.087	−0.047 (0.072)	−1.929
FHZ Standard	0.027 (0.188)	–	0.012 (0.119)	–	0.025 (0.066)	–
Straight Down	−0.009 (0.193)	**−2.798 ***	−0.005 (0.118)	−1.547	−0.001 (0.058)	**−3.667 ***
Straight Up	0.018 (0.180)	−1.964	0.019 (0.115)	**−2.868 ***	0.034 (0.070)	**−3.215 ***
Right Straight	−0.013 (0.206)	−0.226	0.021 (0.167)	−0.365	0.028 (0.069)	−0.921
Right Down	−0.031 (0.204)	**−2.728 ***	−0.011 (0.152)	−0.261	−0.004 (0.075)	**−3.806 ***
Right Up	0.004 (0.179)	−0.904	0.027 (0.127)	**−2.763 ***	0.027 (0.066)	−2.589
Left Straight	−0.016 (0.175)	−0.956	0.026 (0.179)	−0.608	0.025 (0.073)	−1.477
Left Down	−0.009 (0.188)	**−3.041 ***	0.013 (0.171)	−0.504	0.018 (0.070)	−2.311
Left Up	0.007 (0.195)	−0.991	0.028 (0.173)	−1.130	0.033 (0.075)	**−3.945***
PAL Standard	0.021 (0.163)	–	0.042 (0.118)	–	0.042 (0.062)	–
Straight Down	0.029 (0.155)	−1.894	0.028 (0.120)	−2.346	0.032 (0.079)	−1.130
Straight Up	0.006 (0.163)	−2.207	0.054 (0.121)	−0.365	0.027 (0.087)	−1.790
Right Straight	0.037 (0.177)	−0.365	0.040 (0.132)	−0.365	0.034 (0.071)	−1.095
Right Down	0.030 (0.164)	−2.103	0.030 (0.108)	−2.103	0.034 (0.077)	−0.295
Right Up	0.007 (0.165)	−1.721	0.042 (0.116)	−0.400	0.026 (0.089)	−0.504
Left Straight	0.014 (0.147)	−0.156	0.048 (0.119)	−0.261	0.040 (0.066)	−0.261
Left Down	0.023 (0.162)	−1.581	0.045 (0.133)	−0.956	0.041 (0.079)	−0.017
Left Up	−0.004 (0.162)	−2.207	0.055 (0.126)	−1.130	0.026 (0.087)	−0.747

* Bold text with asterisk indicates significance at the Bonferroni-adjusted alpha level of 0.006.

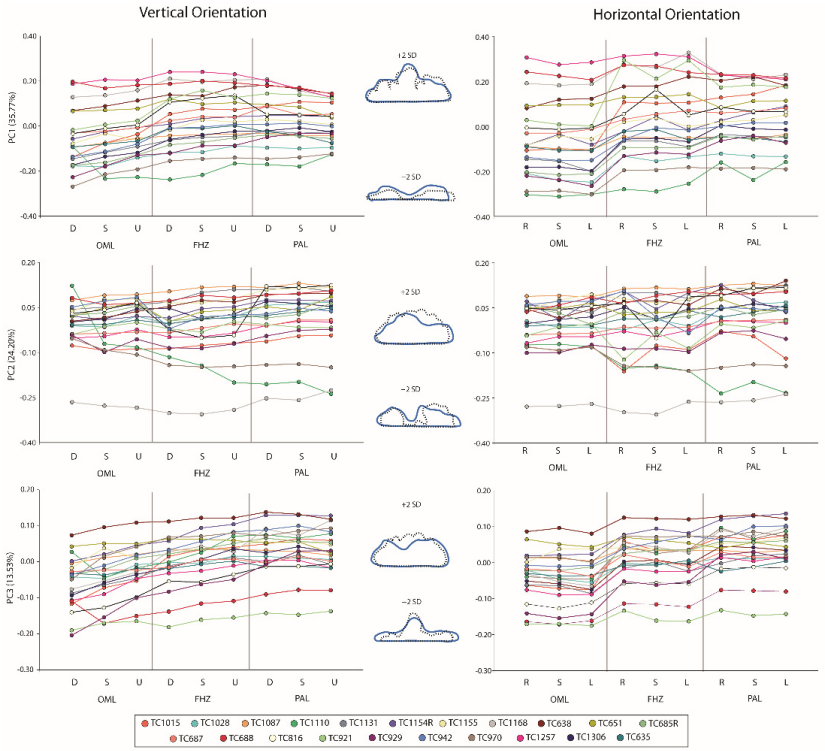


Figure 3. PC1 (top), PC2 (middle), and PC3 (bottom) values plotted against vertical orientations (left) and horizontal orientations (right) for each view; circles with same color scheme represent the same individual across all graphs (legend provided). PC contours also provided, with thick blue lines representing constructed +2 standard deviations (above) and −2 standard deviations (below) relative to the mean; black dashed lines representing actual outlines from an individual near the extremes of the axes. D, down; S, straight; U, up; R, right; L, left.

Table 3. Median and inter-quartile range (IQR) statistics for sinus variables across the three views (Orbitomeatal Line, OML; Frankfort Horizontal Plane, FHZ; and Porion-Alveolar Line, PAL) and varying orientations. Wilcoxon sign rank tests (Z scores and significance) also provided for each orientation versus respective standard view.

Orientation	Area (cm^2)		Height (cm)		Breadth (cm)	
	Median (IQR)	Z	Median (IQR)	Z	Median (IQR)	Z
OML Standard	5.037 (7.077)	–	1.505 (1.287)	–	**5.815 (2.769)**	–
Straight Down	5.251 (7.719)	**−3.146** *	1.394 (1.209)	**−3.181** *	5.783 (2.823)	−1.303
Straight Up	6.681 (7.064)	**−3.597** *	1.638 (1.239)	**−3.806** *	5.862 (2.753)	−0.800
Right Straight	4.867 (7.640)	−0.382	1.553 (1.399)	−0.226	5.846 (2.866)	−0.678
Right Down	5.458 (6.929)	**−2.728** *	1.567 (1.316)	**−3.163** *	5.780 (2.785)	−0.608
Right Up	5.222 (6.370)	**−3.563** *	1.385 (1.247)	**−3.667** *	5.859 (2.864)	−0.417
Left Straight	6.160 (6.903)	−0.896	1.496 (1.204)	−1.717	5.757 (2.754)	−1.304
Left Down	6.332 (7.014)	**−3.041** *	1.678 (1.356)	**−3.250** *	5.763 (2.698)	−1.512
Left Up	5.450 (6.715)	**−3.563** *	1.451 (1.235)	**−3.389** *	5.801 (2.757)	−1.981
FHZ Standard	7.129 (7.775)	–	1.875 (1.347)	–	5.838 (2.742)	–
Straight Down	7.063 (7.337)	**−3.007** *	1.762 (1.342)	**−3.233** *	5.793 (2.731)	−0.463
Straight Up	7.806 (7.385)	**−3.493** *	1.866 (1.292)	**−3.007** *	5.836 (2.734)	−1.565
Right Straight	7.118 (7.689)	−0.504	1.793 (1.428)	−0.574	5.770 (2.807)	−0.205
Right Down	6.956 (7.918)	**−3.389** *	1.739 (1.386)	**−4.015** *	5.788 (2.779)	−0.017
Right Up	7.367 (7.826)	**−3.424** *	2.019 (1.262)	**−2.798** *	5.836 (2.859)	−1.026
Left Straight	7.015 (7.363)	−0.678	1.879 (1.366)	−0.330	5.814 (2.709)	−1.321
Left Down	6.534 (7.482)	**−3.736** *	1.738 (1.391)	**−3.910** *	5.788 (2.743)	−2.312
Left Up	7.525 (7.555)	−2.485	1.965 (1.298)	**−3.245** *	2.709 (2.753)	−1.651
PAL Standard	8.002 (7.883)	–	1.945 (1.196)	–	5.775 (2.708)	–
Straight Down	8.446 (7.853)	−1.095	2.050 (1.348)	**−3.233** *	5.780 (2.753)	−0.417
Straight Up	8.084 (7.420)	−2.033	1.897 (2.169)	**−3.007** *	5.790 (2.729)	−1.363
Right Straight	8.081 (8.182)	−1.303	1.958 (1.237)	−0.574	5.869 (2.840)	−0.037
Right Down	7.999 (7.664)	−0.226	2.041 (1.249)	**−4.015** *	5.854 (2.765)	−0.672
Right Up	7.948 (7.542)	−1.721	1.932 (1.093)	**−2.798** *	5.812 (2.799)	−0.485
Left Straight	8.119 (7.747)	−0.817	1.920 (1.214)	−0.330	5.897 (2.710)	−1.547
Left Down	7.730 (16.219)	−0.261	2.028 (1.303)	**−3.910** *	5.875 (2.730)	−1.095
Left Up	7.741 (7.408)	−2.172	1.930 (1.140)	**−3.245** *	5.770 (2.657)	−0.672

* Bold text with asterisk indicates significance at the Bonferroni-adjusted alpha level of 0.006.

PC Distances

Figure 5 provides histograms of the intra-individual and inter-individual distances for each standard view, while Table 4 provides the descriptive statistics and Mann–Whitney U Test results. For each of the three views, the intra-individual distances were significantly lower than the inter-individual differences (all p-values < 0.005). When comparing the intra-individual distances across the three views, a Kruskal–Wallis test indicated significant differences (test statistic = 118.22; $p < 0.001$). Follow up Mann–Whitney U tests on the intra-individual distances found that the significant differences between all views: OML versus FHZ ($Z = -3.277$; $p = 0.001$); OML versus PAL views ($Z = -10.676$; $p < 0.001$); and FHZ versus PAL views ($Z = -7.252$; $p < 0.001$). Figure 6 provides a boxplot for the intra-individual distances, by individual and view. Notably, the median and range of intra-individual distances are higher among OML views for most individuals compared to the FHZ and PAL views.

Table 4. Median and interquartile range (IQR) values, with Mann–Whitney U results for the multivariate PC intra- and inter-individual distances across the three views: Orbitomeatal Line (OML), Frankfort Horizontal Plane (FHZ), and Porion-Alveolar Line (PAL).

View	Intra-Distances	Inter-Distances	Z
	Median (IQR)	Median (IQR)	
OML	0.053 (0.045)	0.222 (0.144)	**−43.681** *
FHZ	0.048 (0.043)	0.253 (0.145)	**−44.868** *
PAL	0.035 (0.033)	0.225 (0.136)	**−46.183** *

* Bold text with asterisk indicates significance at the 0.05 level.

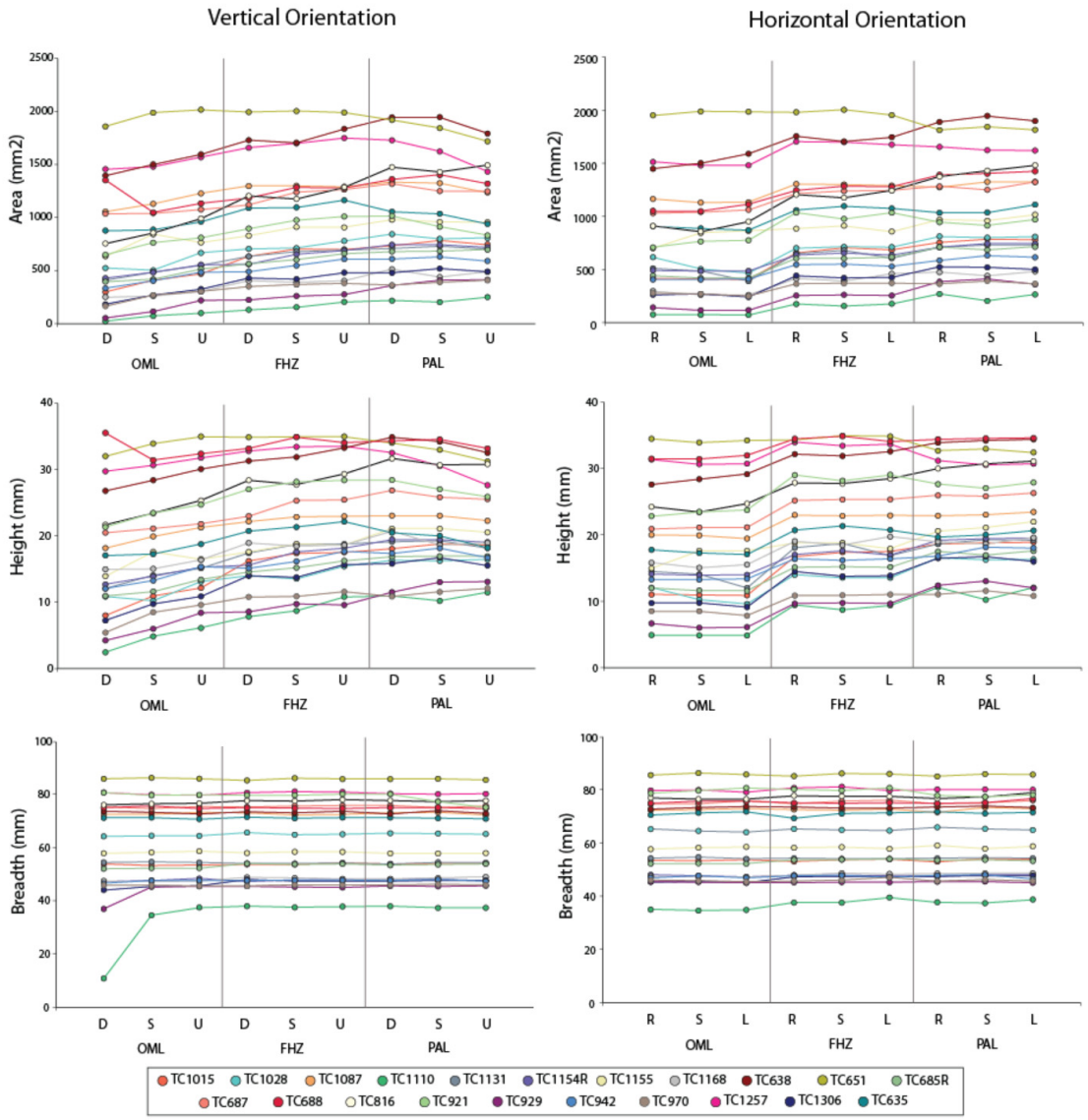


Figure 4. Sinus area (top), height (middle), and breadth (bottom) plotted against vertical orientations (left) and horizontal orientations (right) for each view: Orbitomeatal Line (OML), Frankfort Horizontal Plane (FHZ), and Porion-Alveolar Line (PAL). Circles with same color scheme represent the same individual across all graphs (legend provided). D, down; S, straight; U, up; R, right; L, left.

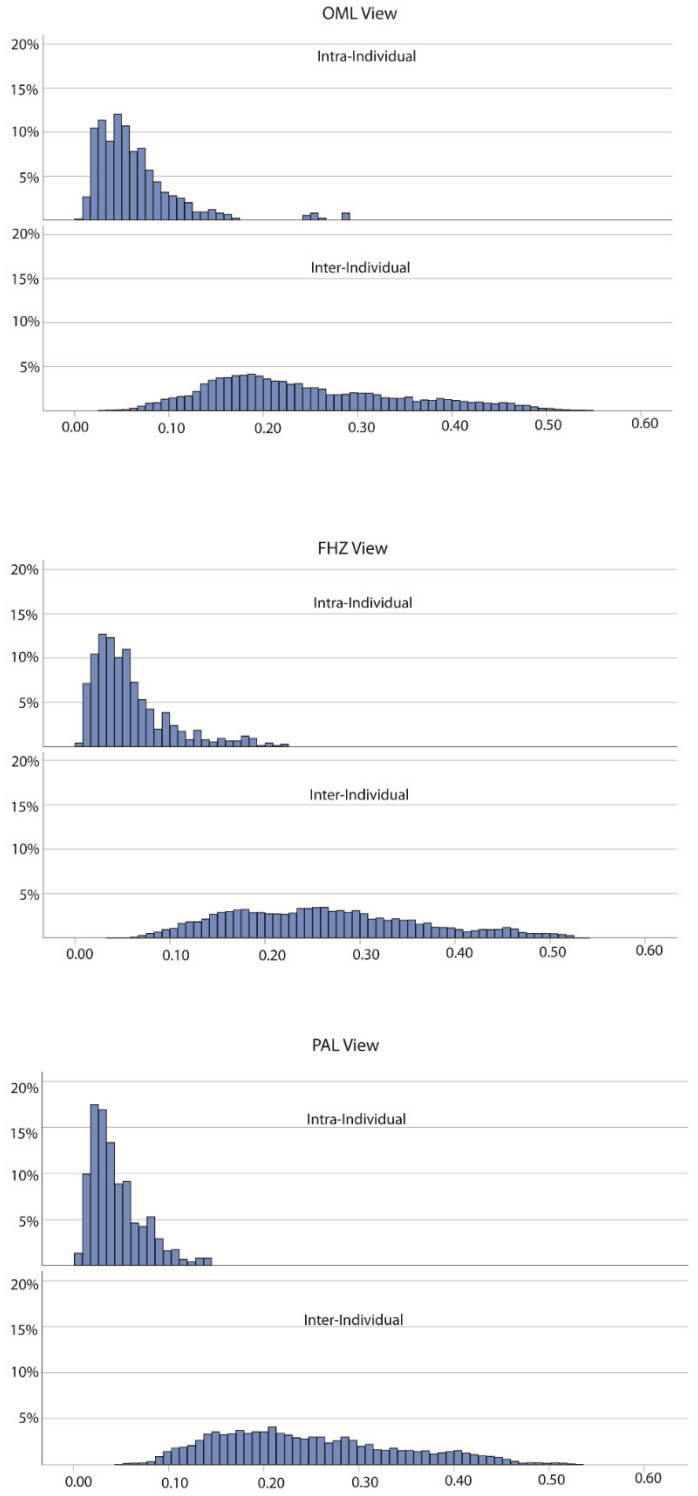


Figure 5. Percentage histogram of multivariate PC inter- and intra-individual distances for Orbitomeatal Line (OML; top); Frankfort Horizontal (FHZ; middle) and Porion-Alveolar Line (PAL; bottom) views.

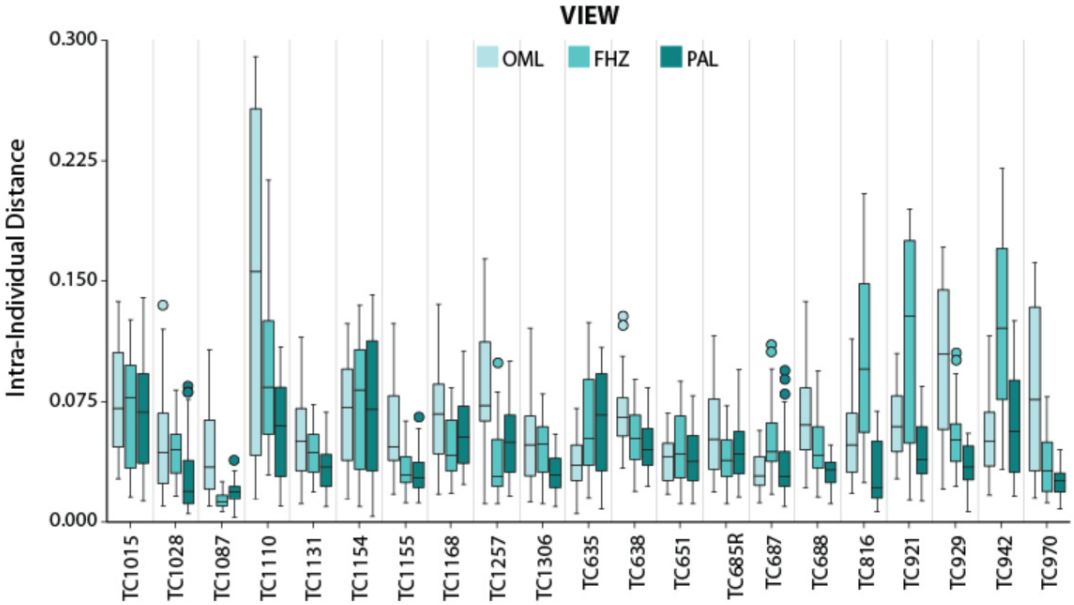


Figure 6. Box and whisker plots illustrating medians and quartiles for intra-individual distances for each individual in the three views: Orbitomeatal Line (OML), Frankfort Horizontal Plane (FHZ), and Porion-Alveolar Line (PAL).

Still, all three views showed high reliability in outline matching. For the PAL and FHZ views, all 189 outlines matched with outlines from their same individual, for a correct match rate of 100%. The OML view had a correct match rate of 98.94%, with only two instances where outlines most closely matched to a different individual. Interestingly, both mismatched instances included the same two individuals, but different paired orientations: TC1154R left-up to TC1155 right-straight and TC1154R straight-down to TC1155 left-straight. Visual assessment of these outlines suggests striking morphological similarities, as seen in Figure 7. Overall, these results suggest that although the 5° deviations returned statistical differences in the Wilcoxon sign rank tests, these deviations are not likely to impact forensic sinus matches in practice, particularly for the PAL and FHZ views.

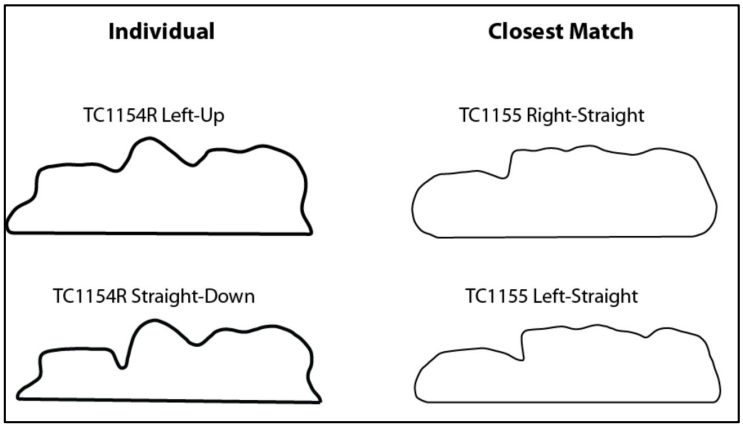


Figure 7. Outlines of the two mismatched individuals (TC1154, TC1155).

4. Discussion

Several studies suggest that when using the frontal sinus as an identification method, postmortem radiographs should be taken as closely as possible to the antemortem orientation [1,19,55,56]. However, a perfect alignment match can be challenging, if not impossible, and small deviations between the radiographic comparisons (e.g., 5°) is highly likely [29,57,58]. The current study investigated how small 5° deviations in vertical, hor-

izontal, and diagonal axes may affect frontal sinus morphology within three clinically relevant views based on the Orbitomeatal Line (OML), Frankfort Horizontal Plane (FHZ), and the Porion-Alveolar Line (PAL).

In terms of overall sinus dimensions, the current study found that sinus breadth remained relatively stable throughout the deviations, while sinus height was more affected by small variations in vertical orientation. This effect is illustrated in Figure 4 (middle left), which shows the progression of change in sinus height from the most inferiorly oriented view (OML straight-down) to the most superiorly oriented view (PAL straight-up). Changes affecting sinus height dimensions will also affect shape variables (PCs) related to sinus height–breadth dynamics (i.e., height relative to breadth). This effect was evident in the Wilcoxon sign rank tests, whereby deviations in vertical orientations resulted in significant differences in PC1 and PC3, both of which capture aspects of sinus height relative to breadth. As further support, horizontal changes without vertical re-alignment (e.g., simply looking left-straight or right-straight) did not present significant differences from the standard views. As discussed further below, these results are important for forensic frontal sinus matching methods that utilize measures of sinus height or variables associated with sinus height (e.g., area, height/breadth indices, or outlines).

When comparing the three standard views, the PAL view showed to be the most stable, with the OML view the least stable, in terms of minor deviations altering sinus morphology. While the PAL view had significant differences in measured height with vertical deviations, all other measured and PC variables appear unaffected. Further evidence for reduced reliability in the OML view comes from the analyses on the Euclidean distances, which showed a higher range of intra-distance variation compared to the other views (Figure 5). As the OML view was the most inferiorly rotated of the three views investigated here, this suggests that more superiorly oriented radiological views are more stable in frontal sinus identifications. This is likely true up until a certain extent, as previous studies have shown that Water's view (involving a superior rotation approximately 10°s higher than our PAL view) can be highly affected by orientation deviations and drastically affects the appearance of craniofacial structures [47]. In terms of morphological stability, frontal sinus dimensions seem to be more stable across varying orientations within more moderate views, such as the PAL and FHZ views utilized here. This is consistent with previous publications [25,26]. In fact, Nikolova et al. [26] found that sinus dimensions in their "Caldwell" view (the same view as our PAL view; see materials/methods) was the least affected by deviations in orientations.

However, previous studies also noted a more drastic alteration of sinus breadth versus height, which is contrary to the results found in the current study. Of several craniofacial measures, Riepert et al. [27] found frontal sinus breadth presented with some of the highest deviations with varying orientations. Nikolova et al. [26] found that any 5° inferior vertical orientations of the cranium from their standard "Caldwell"/PAL position resulted in a significant decrease in sinus breadth, while superior vertical orientations resulted in an increase in breadth. This likely relates to the fact that most sinuses are widest near their inferior base and the inferior border was determined at the superior orbital margin. As such, a superior tilting of the cranium would result in the base of the sinus being more prominent, while inferior tilting of the cranium would result in the base of the sinus dipping below the superior orbital line. Silva et al. [25] also found that 10° vertical deviations from their standard "Caldwell"/PAL position resulted in narrowing of breadths, likely due to the lateral edges of the sinus being lost from view in either direction.

There could be several reasons why the current study did not also find significant differences in breadth. First, it should also be noted that the Silva et al. [25] study focused on relatively large degrees of variations (10°); there is a possibility that significant differences in sinus breadth would not have been found in smaller degrees of orientation, such as the 5°s measured here. Additionally, crania were manually repositioned to obtain each varying degree, which could have introduced an additional potential source of error [29]. As another consideration, both Silva et al. and Nikolova et al. utilized radiographic

images, which incorporate measures of distortion inherent to radiographs, including issues of superimposition and magnification, that were not investigated here with CT-derived models. Yanagisaw and Smith [24] (p. 112) note that "the posterior tilt of the head ... causes some distortion of the frontal sinuses because their vertical axes are not parallel to the film and the space between the frontal sinuses and the film is considerable." It could very well be that these radiographic sources of visualization error more greatly affect sinus breadth versus height, which the current study did not capture (see Section 4.2) also see [27,58].

4.1. Effect on Sinus Identification

Still, despite significant differences between several orientations and views found in the current study, these deviations do not seem to largely affect potential identification as assessed by outline analyses (e.g., the Christensen method [15,24]). For all three views, Mann–Whitney U tests indicate that the intra-individual differences were significantly lower than the inter-individual differences. All three views also had high instances of true-positive matches, with OML at 98%, FHZ at 100%, and PAL at 100%. These results are similar to Christensen [15], who found that while there was some overlap where individual outlines most closely matched another individual, such occurrences were rare. Overall, this suggests that in most cases—regardless of varying 5° orientations or views—an individual's outlines more closely resemble each other than outlines of other individuals. This is likely due to the already high inter-individualistic aspect of sinus morphology [27], which supersedes more subtle differences related to orientation. However, given that two outlines for the OML view did incorrectly match with another individual, caution is warranted when applying such quantitative methods of sinus identification. Although the outlines that erroneously matched in this study were strikingly similar (Figure 7), a simple visual assessment could likely distinguish the two. This suggests that current quantitative methods are not yet capable of distinguishing subtle differences or performing the complex interpretations undertaken by human observation. Thus, while there is a push to move towards more quantitative and objective methods in the forensic sciences (particularly in the U.S. since the Daubert guidelines [59,60] and 2009 National Academy of Sciences Report [61]), such methods may be more susceptible to noise and minor deviations.

Certain frontal sinus morphologies may also be more unique than others, and additional analyses are required to assess whether certain sinus variables (e.g., size, degree of complexity, etc.) are more prone to deviation issues or mismatches in larger and more diverse outline samples. There is already some indication that frontal sinus size affects the reliability of identification rates. As previously noted by Christensen [28] and Smith et al. [62], smaller sinuses are typically less complex (e.g., in terms of arcade number) and, thus, less diagnostic for identification purposes. Further, even slight vertical and/or horizontal variations could cause a smaller sinus to be partially, or even entirely, eliminated from view. Along these lines, the current study points to caution warranted when using identification methods on discontinuous sinuses and/or sinuses that have smaller lobes or lower arcades near the superior orbital border. In such cases, even small 5° variations may drastically alter the shape of the sinus. In the current study, this is best seen in Figure 4 (bottom, left), which shows the case of a single individual whose breadth was drastically smaller in the OML straight-down versus other views. Figure 8 investigates the OML outlines of this individual further. Note how, when in standard OML view, the individual outline presents with two distinct sinus lobes, with the anatomically right sinus being smaller (indicated by the large arrow) than the left; the left lobe also possesses a small arcade on its lateral edge (small arrow). As the individual is re-oriented inferiorly, the right lobe and smaller arcade completely disappear from view. This change would result in drastically different PC scores, particularly in terms of PC1 and PC3, which both track height-breadth dynamics. In a real-case scenario, if an investigator only had two images for this individual (e.g., straight-down and standard, see Figure 8), a true positive identification could be missed.

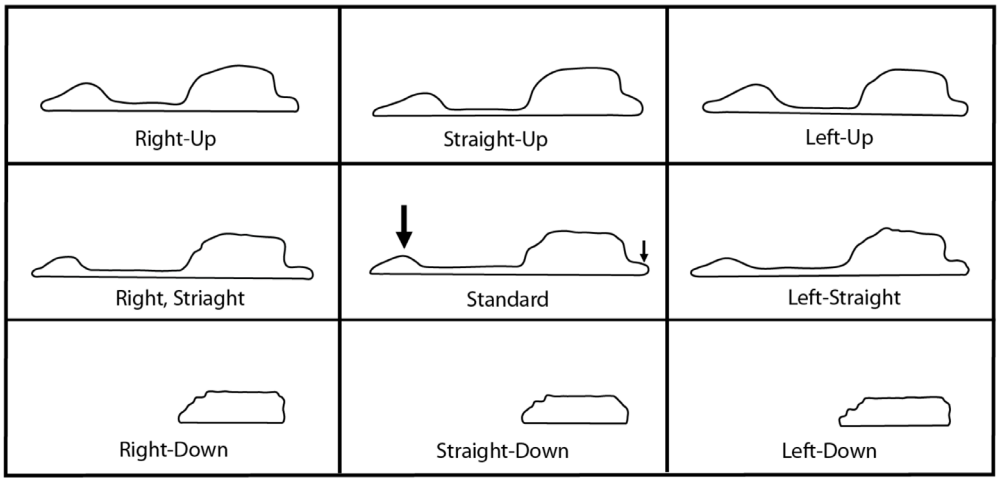


Figure 8. Example of individual (TC1110) outlines with relatively high intra-distances in the Orbitomeatal Line (OML) view (0.25–0.30) (see Figure 8 histogram); note the loss of the anatomical-right lobe and tail (large arrow), as well as the loss of the smaller anatomically left-sided arcade (small arrow), in the inferiorly oriented (Down) views. This drastically alters shape outline from two almost discontinuous lobes to a single plateau-like lobe. Outlines represent the normalized elliptical Fourier coefficients (based on 20-harmonics and aligned by the first harmonic).

4.2. Limitations and Future Directions

While this study provides a preliminary understanding on how deviations in radiographic orientations can affect frontal sinus identification methods, there are several considerations for future studies. First, although we modeled this study based on the EFA method developed by Christensen [15,16,28,54], it was slightly modified from original analysis. Both studies utilized EFA, a standard method of comparing closed outlines, but the collection and analyses of these outlines differed. Using conventional radiographic images, Christensen manually traced outlines on acetate paper, then digitally converted the images into x, y coordinates. The act of manual tracing has potential for error due to differences in tracing the contours. The incorporation of 3D models here (see Materials and Methods) allowed the automatic capturing of outline shape without the potential effect (no matter how minimal) of manual tracing error. Further, unlike Christensen who utilized the x, y coefficients of the outline harmonics as their primary variable, we conducted a PCA to obtain PCs as our shape variables. The use of PCA has the advantage of simplifying the dataset, which are easier to analyze and interpret. In both approaches, the overall results were the same: the use of EFA on outline shape analyses provides a relatively robust method for frontal sinus outline identification, at least for larger continuous sinuses (see above). Still, additional studies further testing the inter- and intra-reliability of this method across a wide range of sinus sizes and observers is necessary to fully validate this method. Along these lines, additional studies focusing on relationships between frontal sinus patterns and body size and/or cranial–facial morphology would be beneficial. Indeed, while several studies attempt to discern such relationships across diverse human populations, such studies are conflicting and no consensus of the underlying factors explaining sinus morphology have been reached (for more discussion, see [34,36,38]). Further, the direct implications (if any) of these variations on frontal sinus morphology, particularly outline shape, to forensic identifications are lacking in the literature.

Of importance to consider here is that all frontal sinus identification methods, whether based on coding, metric, visualization, or outlines, take the same variables (e.g., height vs. breadth, arcade/scallop number and presentation, presence/absence of sinus lobes) into consideration when attempting to corroborate or negate a potential match. While we focused on a single outline method, the other methods could also be affected by the varying orientations altering sinus shape described here. For example, a coding method incorporating the number of arcades or presence/absence of sinus lobes (e.g., [7,22]) would

understandably be affected if a lobe or arcade became unviewable with certain orientations. Additional testing across wider ranges of identification methods and among more diverse samples is needed to determine if any single method is more robust to alteration in orientations compared to others.

Finally, the primary limitation of this study is that it does not directly emulate real-life scenarios, largely due to the incorporation of CT-derived models versus radiographs. While using 2D images of the segmented sinuses emulates the 2D view obtained in traditional radiographs, the CT-derived sinuses would not have been affected by radiographic-specific parameters, such as radiographic quality, magnification, distortion or human error in degree placement. Overall, CTs have several benefits over radiographs: they allow a clearer, 3D view of the sinuses, can be oriented in any direction, and digitally derived models can be modified to showcase soft tissue, if needed, or not [63]. Further, emerging technological advantages in the clinical sector will likely result in investigators being presented with more antemortem images from CT scans [64]. Owing to this, several studies attempt to create frontal sinus identification methods based on CT scans and digitally derived sinus models specifically [8,9,17,65–72]. However, in terms of postmortem images, all investigators may not have the time, resources, or experience to obtain and evaluate CT scans, let alone go through the process of creating frontal sinus models for identification purposes. Access to such technology may vary depending on geographic regions as well. While postmortem virtual autopsies utilizing CT scans (sometimes referred to as virtopsies or postmortem CTs) are relatively common among European countries [73], their incorporation in the United States is lagging. In fact, a relatively recent article from 2018 indicated that only four U.S. agencies have CT equipment available for regular use [74].

Although postmortem analog radiographs may still be more likely (at least in certain regions), we avoided two potential sources of error by utilizing digital frontal sinus models versus traditional radiographs in the current study. First, we more accurately and precisely positioned the cranium using digital means, avoiding error introduced by manual repositioning crania on the X-ray tables [26,29]. Second, since the act of digitally or manually tracing outlines may impose additional error, we obtained 2D images of the models to automatically digitize frontal sinus outlines. Both steps allowed us to directly test the actual effect of varying orientation on this method, without additional sources of error and/or bias. However, this also means that the major sources of error introduced in varying orientations on frontal sinus morphology—issues of distortion, magnification, and superimposition of radiographs—could not be taken into consideration here. These effects were assessed by Nikolova et al. [26] who directly tested how changes in radiographic images taken from an industrial μCT scanner distorted actual linear measures. By directly comparing radiographic linear measures to a virtual frontal sinus endocast (i.e., model), they found breadth is more distorted than height dimensions. However, while they used an industrial μCT scanner, which has a fixed X-ray tube and flat panel detector similar to conventional radiography, it is unclear whether the beam passed through the crania from a posterior–anterior or anterior–posterior direction—an important distinction, as the differing orientation of the beam will pass through different layers of superimposed structures, varying the effects of distortion. In the clinical setting, the radiographic beams are typically aligned posteriorly–anteriorly through the cranium, which avoids direct radiation that could harm the patient's orbital contents. But, there is further distortion of the sinus morphology as the beam travels through the present soft tissue, such as brain matter [75]. Future studies directly comparing frontal sinus identification methods between mixed modalities (e.g., CT scans vs. traditional radiographs) and with the presence/absence of soft tissue would be informative.

4.3. Recommendations on Sinus Identifications

While there will hopefully be a move to include more advanced imagining technology at medicolegal agencies globally, we offer several recommendations when using the frontal sinus as an identification method with analog radiographs. Firstly, it is obvious that

practitioners should aim to orient the cranium as close as possible to the antemortem comparative image. Besides the position of the skull, this may also involve matching the direction of X-ray (e.g., anterior–posterior or posterior–anterior) and the angle at which the X-ray enters the skull and encounters the film. For example, in true Caldwell position while the head is oriented horizontally to the film, the X-ray beam is oriented at a 15–20° angle. With this in mind, forensic anthropologists conducting radiographic comparisons for personal identification should ideally have training in radiographic techniques. An increase in collaborations between clinicians, radiographic technicians, and medicolegal practitioners can ensure a better understanding of radiological practices and consensus on terminology.

In terms of application, the results presented here can help practitioners differentiate explainable from unexplainable differences when conducting radiographic comparisons of frontal sinus morphologies. Although this study focused on one sinus outline method of identification, understanding the sinus shape variations that are expected with slight orientation differences can be used to better interpret the results of other sinus identification methods. Differences in vertical orientation of the crania can be expected to affect sinus height and/or the presence of smaller lobes/arcades, particularly those near the supraorbital border. In such cases, identification methods that rely on height measurements or counts of lobes/arcades should be avoided. If applied, a visual comparison should be used to confirm results, taking into account these expected changes with orientation, to ensure that a correct identity is not erroneously excluded due to these methodological errors. Keep in mind that such variation may be more drastic in more extreme views, such as Water's view (common in clinical settings) and/or the OML view assessed here. Ultimately, although there is a push in forensics towards more quantitative methods such as EFA/outline analyses, coding methods, and metric analyses for the frontal sinus, these more objective methods may be more sensitive to slight deviations in the capture of the radiographic frontal sinus. Until more robust methods of quantitatively describing the frontal sinus morphology are developed or the use of CT technology becomes more commonplace across medicolegal agencies globally, visual comparisons, albeit more subjective, are likely more capable of interpreting such explainable differences between radiographs.

5. Conclusions

Overall, the current study found that the EFA outline method for identification developed by Christensen [15,16,28] is relatively robust to small 5° variations in orientation. However, in conjunction with previous studies, it is evident that reliability of frontal sinus identification methods is largely contingent on the view being imitated, the directionality of the deviation, and actual sinus morphology. Furthermore, based on our results, Christensen's EFA outline method would appear to be most reliable on larger sinuses, particularly those that are superior inferiorly tall and medio-laterally wide without discontinuous lobes (i.e., the right and left sinuses are touching), although additional testing is needed to validate this method. For any method (e.g., outline, coding, and linear), caution is warranted when attempting to identify individuals with small sinuses, particularly sinuses possessing smaller discontinuous lobes.

In terms of clinical views, small degrees of varying orientation within an intermediate range of standard clinical views (e.g., FHZ, PAL, and true Caldwell view) would be more reliable compared to other views. Caution is highly warranted if attempting to match antemortem radiographs taken in more extreme vertical orientations, including those based on the Orbitomeatal Line (e.g., posterior–anterior frontal view) and/or or Water's view, as small deviations from that standard view can have more drastic effects on the presentation of sinus morphology. Finally, the results of this CT-based study should be considered limited, as additional sources of distortion common in traditional radiographs (e.g., angulation, magnification, presence of soft tissue, and superimposition) will likely have greater alterations to sinus morphology, particularly breadth, than that presented

here. The results of this study assist practitioners in better understanding and interpreting explainable and unexplainable differences between radiographs.

Author Contributions: Conceptualization, L.N.B. and H.M.G.; methodology, L.N.B. and A.R.; formal analysis, L.N.B. and H.M.G.; investigation, A.R., L.N.B. and H.M.G.; resources, L.N.B.; data curation, L.N.B.; writing—original draft preparation, L.N.B. and H.M.G.; writing—review and editing, A.R.; supervision, L.N.B.; project administration, L.N.B.; funding acquisition, L.N.B. and H.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by Des Moines University Iowa Osteopathic Education and Research R&G (Award No. 03-19-01) and Award No. 2020-75-CX-0013 granted 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 publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Institutional Review Board Statement: The research in this study was conducted in compliance with the Declaration of Helsinki and federal regulations for the protection of human subjects in research at 45 CFR 46. This research involved neither living individuals nor data from living humans, and all data involved are unidentifiable and, thus, per the aforementioned regulations, do not require institutional review.

Informed Consent Statement: Not applicable.

Data Availability Statement: CT scans publicly available online. Other data available on request from the authors.

Acknowledgments: The authors would like to thank L Copes for making CT scans freely available on their website, J Campbell for providing statistical advice, and D Elling for providing macros to facilitate data collection.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CT	computed tomographic
EFA	elliptical Fourier analysis
FHZ	Frankfort Horizontal Plane
OML	Orbitomeatal Line
PA	posterior–anterior
PAL	Porion-Alveolar Line
PCA	principal components analysis
TC	Terry Collection

References

1. Culbert, W.L.; Law, F.L. Identification by comparison of roentgenograms of nasal accessory sinuses and mastoid processes. *JAMA* **1927**, *88*, 1634–1636. [CrossRef]
2. Mayer, J. Identification by sinus prints. *VA Med. Mon.* **1935**, *62*, 517–519.
3. Schuller, A. A note on the identification of skulls by X-ray pictures of the frontal sinuses. *Med. J. Aust.* **1943**, *1*, 554–556. [CrossRef]
4. Murphy, W.A.; Spruill, F.G.; Gantner, G.E. Radiologic identification of unknown human remains. *J. Forensic Sci.* **1980**, *25*, 727–735. [CrossRef] [PubMed]
5. Murphy, W.; Gantner, G. Radiologic examination of anatomic parts and skeletonized remains. *J. Forensic Sci.* **1982**, *25*, 727–735. [CrossRef]
6. Ubelaker, D.H. Positive identification from radiographic comparison of frontal sinus patterns. In *Human Identification—Case Studies in Forensic Anthropology*; Rathbun, T.A., Buikstra, J.E., Eds.; CC Thomas: Springfield, IL, USA, 1984; pp. 399–411.
7. Yoshino, M.; Miyasaka, S.; Sato, H.; Seta, S. Classification system of frontal sinus patterns by radiography. Its application to identification of unknown skeletal remains. *Forensic Sci. Int.* **1987**, *34*, 289–299. [CrossRef]
8. Reichs, K.J.; Dorion, R.B.J. The use of computerized axial tomography scans in the comparison of frontal sinus configurations. *Can. Soc. Forensic Sci. J.* **1992**, *25*, 1–16. [CrossRef]
9. Reichs, K.J. Quantified comparison of frontal sinus patterns by means of computed tomography. *Forensic Sci. Int.* **1993**, *61*, 141–168. [CrossRef]

10. Haglund, W.; Fligner, C. Confirmation of human identification using computerized tomography (CT). *J. Forensic Sci.* **1993**, *38*, 708–712. [CrossRef] [PubMed]
11. Owsley, D. Identification of the fragmentary, burned remains of two us journalists seven years after their disappearance in Guatemala. *J. Forensic Sci.* **1993**, *38*, 1372–1383. [CrossRef] [PubMed]
12. Quatrehomme, G.; Fronty, P.; Sapanet, M.; Grevin, G.; Bailet, P.; Ollier, A. Identification by frontal sinus pattern in forensic anthropology. *Forensic Sc.i Int.* **1996**, *83*, 147–153. [CrossRef]
13. Nambiar, P.; Naidu, M.D.; Subramaniam, K. Anatomical variability of the frontal sinuses and their application in forensic identification. *Clin. Anat.* **1999**, *12*, 16–19. [CrossRef]
14. Smith, D.R.; Limbird, K.G.; Hoffman, J.M. Identification of human skeletal remains by comparison of bony details of the cranium using computerized tomographic (CT) scans. *J. Forensic Sci.* **2002**, *47*, 937–939. [CrossRef]
15. Christensen, A.M. Assessing the variation in individual frontal sinus outlines. *Am. J. Phys. Anthropol.* **2004**, *127*, 291–295. [CrossRef] [PubMed]
16. Christensen, A.M. Testing the reliability of frontal sinuses in positive identification. *J. Forensic Sci.* **2005**, *50*, JFS2004145-5. [CrossRef]
17. Tatlisumak, E.; Ovali, G.Y.; Aslan, A.; Asirdizer, M.; Zeyfeoglu, Y.; Tarhan, S. Identification of unknown bodies by using CT images of frontal sinus. *Forensic Sci. Int.* **2007**, *166*, 42–48. [CrossRef] [PubMed]
18. Campobasso, C.P.; Dell'Erba, A.S.; Belviso, M.; Di Vella, G. Craniofacial idendentification by comparison of antemortem and postmoretm radiographs: Two case reports dealing with burnt bodies. *Am. J. Forensic Med.* **2007**, *28*, 182–186. [CrossRef]
19. Asherson, N. *Identification by Frontal Sinus Prints*; HK Lewis: London, UK, 1965.
20. Cryer, M. Some variations in the frontal sinuses. *JAMA* **1907**, *48*, 284–289. [CrossRef]
21. Ribeiro, F.A.Q. Standardized measurements of radiographic films of the frontal sinuses: An aid to identifying unknown persons. *Ear Nose Throat J.* **2000**, *79*, 26–33. [CrossRef]
22. Cameriere, R.; Scendoni, R.; Lin, Z.; Milani, C.; Palacio, L.A.; Turiello, M.; Ferrante, L. Analysis of frontal sinuses for personal identificaiton in a Chinese sample using a new code number. *J. Forensic Sci.* **2020**, *65*, 46–51. [CrossRef] [PubMed]
23. Cox, M.; Malcolm, M.; Fairgrieve, S.I. A new digital method for the objective comparison of frontal sinuses for identification. *J. Forensic Sci.* **2009**, *54*, 761–772. [CrossRef]
24. Yanagisawa, E.; Smith, H.M. Radiographic anatomy of the paranasal sinuses iv. Caldwell view. *Arch. Otolaryngol.* **1968**, *87*, 109–120. [CrossRef]
25. Silva, R.F.; Vaz, C.G.; Domiciano, M.L.; Franco, A.; Carla Ap, B.D.C.; do Prado, M.M. Radiographic alterations of the frontal sinus morphology according to variations of the vertical angle in posterior-anterior radiographs of the skull. *Acta Sci. Health Sci.* **2014**, *36*, 113–117. [CrossRef]
26. Nikolova, S.; Toneva, D.; Georgiev, I.; Dandov, A.; Lazarov, N. Morphometric analysis of the frontal sinus: Application of industrial digital radiography and virtual endocast. *J. Forensic Radiol. Imaging* **2018**, *12*, 31–39. [CrossRef]
27. Riepert, T.; Ulmcke, D.; Schweden, F.; Nafe, B. Identification of unknown dead bodies by x-ray image comparison of the skull using the X-ray simulation program FoXSIS. *Forensic Sci. Int.* **2001**, *117*, 89–98. [CrossRef]
28. Christensen, A.M. An Empirical Examination of Frontal Sinus Outline Variability Using Elliptic Fourier Analysis: Implications for Identification, Standardization, and Legal Admissibility. Ph.D. Thesis, University of Tennessee, Knoxville, TN, USA, 2003.
29. Hashim, N.; Hemalatha, N.; Thangaraj, K.; Kareem, A.; Ahmed, A.; Hassan, N.F.; Jayaprakash, P.T. Practical relevance of prescribing superimposition for determing a frontal sinus pattern match. *Forensic Sci. Int.* **2015**, *253*, 137.e1–137.e7. [CrossRef]
30. Ubelaker, D.H.; Shamlou, A.; Kunkle, A. Contributions of forensic anthropology to positive scientific identification: A critical review. *Forensic Sci. Res.* **2018**, *4*, 45–50. [CrossRef]
31. Hunt, D.R.; Albanese, J. History and demographic composition of the Robert J. Terry anatomical collection. *Am. J. Phys. Anthropol.* **2005**, *127*, 406–417. [CrossRef]
32. Copes, L.E. Comparative and Experimental Investigations of Cranial Robusticity in Mid-Pleistocene Hominins. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2012.
33. Copes, L.E. CT Scans of Modern Homo Sapiens. 2012. Available online: https://www.lynncopes.com/human-ct-scans.html (accessed on 1 December 2021).
34. Buck, L.T.; Stringer, C.B.; MacLarnon, A.M.; Rae, T.C. Variation in paranasal pneumatisation between mid-Late Pleistocene hominins. *Bull. Mémoires Société d'Anthropologie Paris* **2019**, *31*, 14–33. [CrossRef]
35. Sardi, M.L.; Joosten, G.G.; Pandiani, C.D.; Gould, M.M.; Anzelmo, M.; Ventrice, F. Frontal sinus ontogeny and covariation with bone structures in a modern human population. *J. Morphol.* **2018**, *279*, 871–882. [CrossRef] [PubMed]
36. Butaric, L.N.; Buck, L.T.; Balzeau, A.; Grine, F.E. The paranasal sinuses of the Hofmeyr cranium. In *Hofmeyr—A Late Pleistocene Human Skull from South Africa*; Grine, F.E., Ed.; Springer: Dordrecht, the Netherlands, in press.
37. Stalling, D.; Westerhoff, M.; Hege, H.-C. Amira: A highly interactive system for visual data analysis. In *The Visualization Handbook*; Hansen, C.D., Johnson, C.R., Eds.; Elsevier Butterworth–Heinemann: Burlington, MA, USA, 2005; pp. 749–767.
38. Butaric, L.N.; Stansfield, E.; Vasilyev, A.Y.; Vasilyev, S. CT-based descriptions of the paranasal complex of Sungir-1, an Upper Paleolithic European. *PaleoAnthorpology* **2019**, *2019*, 389–399.
39. Oh, S.; Ahn, J.; Nam, K.U.; Paeng, J.Y.; Hong, J. Frankfort horizontal plane is an appropriate three-dimensinal reference in the evaluation of clinical and skeletal cant. *J. Korean Assoc. Oral Maxillofac. Surg.* **2013**, *39*, 71–76. [CrossRef]

40. Gulsen, A.; Sibar, S.; Ozmen, S. Orthognathic treatment of facial asymmetry due to temporomandibular joint ankylosis. *Arch. Plast. Surg.* **2018**, *45*, 74–79. [CrossRef]
41. Chen, K.; Xiao, D.; Abotaleb, B.; Chen, H.; Li, Y.; Zhu, S. Accuracy of virtual surgical planning in treatment of temporomandibular joint ankylosis using distraction osteogenesis: Comparison of planned and actual results. *J. Oral Maxillofac. Surg.* **2018**, *76*, 2422.e1–2422.e20. [CrossRef]
42. Zebeib, A.M.; Naini, F.B. Variability of the inclination of anatomic horizontal reference planes of the craniofacial complex in relation to the true horizontal line in orthognathic patients. *Am. J. Orthod. Dentofac. Orthop.* **2014**, *146*, 740–747. [CrossRef]
43. Cameriere, R.; Ferrante, L.; Molleson, T.; Brown, B. Frontal sinus accuracy in identification as measured by false positives in kin groups. *J. Forensic Sci.* **2008**, *53*, 1280–1282. [CrossRef]
44. Buyuk, S.K.; Karaman, A.; Yasa, Y. Association between frontal sinus morphology and craniofacial parameters: A forensic view. *J. Forensic Leg. Med.* **2017**, *49*, 20–23. [CrossRef] [PubMed]
45. Zhao, H.; Li, Y.; Xue, H.; Deng, Z.H.; Liang, W.B.; Zhang, L. Morphological analysis of three-dimensionally reconstructed frontal sinuses from chinese han population using computed tomography. *Int. J. Leg. Med.* **2021**, *135*, 1015–1023. [CrossRef] [PubMed]
46. Gasperini, G.; Cruz, R.L.; de Sá, R.T.; Cardoso, L.L.; de Cavalho, I.B. Assessment of the relationship of the Frankfort horizontal plane and the orbitomeatal line with attainment of the natural head position. *J. Diag. Treat. Oral Maxillofac. Pathol.* **2017**, *1*, 156–163. [CrossRef]
47. Merrell, R.A.; Yanagisawa, E. Radiographic anatomy of the paranasal sinuses: I. Water's view. *Arch. Otolaryngol.* **1968**, *87*, 88–99. [CrossRef] [PubMed]
48. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef]
49. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to Imagej: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef]
50. Caple, J.; Byrd, J.; Stephan, C.N. Elliptical Fourier analysis: Fundamentals, applications, and value for forensic anthropology. *Int. J. Leg. Med.* **2017**, *131*, 1675–1690. [CrossRef] [PubMed]
51. Iwata, H.; Ukai, Y. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic fourier descriptors. *J. Hered.* **2002**, *93*, 384–385. [CrossRef] [PubMed]
52. IBM Corporation. *IBM SPSS Statistics for Windows, Version 28.0.*; IBM Corp.: Armonk, NY, USA, 2021.
53. Rosenberg, M.S. Passage: Pattern analysis, Spatial Statistics, and Geographic Exegesis, Version 2. 2008. Available online: www.passagesoftware.net (accessed on 31 December 2021).
54. Christensen, A.M.; Hatch, G.M. Advances in the use of frontal sinuses for human identification. In *New Perspectives in Forensic Human Skeletal Identification*; Latham, K., Bartelink, E., Finnegan, M., Eds.; Academic Press: London, UK, 2018; pp. 227–240.
55. Kirk, J.N.; Wood, R.E.; Goldstein, M. Skeletal identification using the frontal sinus region: A retrospective study of 39 cases. *J. Forensic Sci.* **2002**, *47*, 318–323. [CrossRef]
56. Silva, R.F.; Rodrigues, L.G.; Manica, S.; Franco, R.P.A.V.; Franco, A. Human identification established by the analysis of frontal sinus seen in anteroposterior skull radiographs using the mento-naso technique. *RBOL—Rev. Bras. Odontol. Leg.* **2019**, *6*, 62–66. [CrossRef]
57. Major, P.W.; Johnson, D.E.; Hesse, K.L.; Glover, K.E. Effect of head orientation on posterior anterior cephalometric landmark identification. *Angle Orthod.* **1996**, *66*, 51–60.
58. Malkoc, S.; Sari, Z.; Usumez, S.; Koyuturk, A.E. The effect of head rotation on cephalometric radiographs. *Eur. J. Orthodont.* **2005**, *27*, 315–321. [CrossRef]
59. Christensen, A.M. The impact of *daubert*: Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *J. Forensic Sci.* **2004**, *49*, PAPER ID JFS2003185. [CrossRef]
60. Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. Available online: https://embryo.asu.edu/pages/daubert-v-merrell-dow-pharmaceuticals-inc-1993 (accessed on 21 November 2021).
61. President's Committee of Advisors on Science and Technology. *Report to the President on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*; Executive Office of the President: Washington, DC, USA, 2016.
62. Smith, V.A.; Christensen, A.M.; Myers, S.W. The reliability of visually comparing small frontal sinuses. *J. Forensic Sci.* **2010**, *55*, 1413–1415. [CrossRef]
63. Butaric, L.N.; Jones, G.C.; Garvin, H.M. Revisiting global patterns of frontal sinus aplasia utilizing computed tomography. *Forensic Sci. Int.* **2020**, *315*, 110458. [CrossRef]
64. Viner, M. Overview of advances in forensic radiological methods of human identification. In *New Perspectives in Forensic Human Skeletal Identification*; Latham, K.B., Finnegan, M., Eds.; Academic Press: London, UK, 2018; pp. 217–226.
65. Choi, I.G.G.; Duailibi-Neto, E.F.; Beaini, T.L.; da Silva, R.L.B.; Chilvarguer, I. The frontal sinus cavity exhibits sexual dimorphism in 3d cone-beam ct images and can be used for sex determination. *J. Forensic Sci.* **2018**, *63*, 692–698. [CrossRef] [PubMed]
66. Uthman, A.T.; Al-Rawi, N.H.; Al-Naaimi, A.S.; Tawfeeq, A.S.; Suhail, E.H. Evaluation of frontal sinus and skull measurements using spiral ct scanning: An aid in unknown person identification. *Forensic Sci. Int.* **2010**, *197*, 124.e1–124.e7. [CrossRef]
67. Zhang, H.X.; Yang, C.B.; Guo, L.; Deng, Z.H. Forensic identification by coding of frontal sinus characteristics of CT. *Fa Yi Xue Za Zhi* **2011**, *27*, 13–16.

68. Gach, P.; Tuchtan-Torrents, L.; Delteil, C.; Adalian, P.; Piercecchi, M.D.; Ebert, L.C.; Gorincour, G. Virtual reconstruction of paranasal sinuses from ct data: A feasibility study for forensic application. *Diagn. Interv. Imaging* **2019**, *100*, 163–168. [CrossRef]
69. Brun, C.N.; Christensen, A.M.; Kravarski, M.; Gorincour, G.; Schweitzer, W.; Thali, M.J.; Gascho, D.; Hatch, G.M.; Ruder, T.D. Comparative radiologic identification with standardized single CT images of the paranasal sinuses— evaluation of inter-rater reliability. *Forensic Sci. Int.* **2017**, *280*, 81–86. [CrossRef]
70. De Souza, L.A., Jr.; Marana, A.N.; Weber, S.A.T. Automatic frontal sinus recognition in computed tomography images for person identification. *Forensic Sci. Int.* **2018**, *286*, 252–264. [CrossRef]
71. Kim, D.; Lee, U.; Park, S.; Kwak, D.; Han, S. Identification using frontal sinus by three-dimensional reconstruction from computed tomography. *J. Forensic Sci.* **2013**, *58*, 5–12. [CrossRef] [PubMed]
72. Li, Y.; Xu, C.; Yu, D.; Xiong, T.; Zhao, H.; Xue, H.; Liang, W.B.; Deng, Z.H.; Zhang, L. Computer-aided superimposition of the frontal sinus via 3D reconstruction for comparative forensic identification. *Int. J. Leg. Med.* **2021**, *135*, 1993–2001. [CrossRef]
73. Grabherr, S.; Baumann, P.; Minoiu, C.; Fahrni, S.; Mangin, P. Post-mortem imaging in forensic investigations: Current utility, limitations, and ongoing developments. *Res. Rep. Forensic Med. Sci.* **2016**, *6*, 25–37. [CrossRef]
74. Weiss, D.; McLedo-Henning, D.; Waltke, H. Using advanced imaging technologies to enhance autopsy practices. *NIJ J.* **2018**, *279*, 27–33.
75. Van Vlijmen, O.J.; Maal, T.J.; Bergé, S.J.; Bronkhorst, E.M.; Katsaros, C.; Kuijpers-Jagtman, A.M. A comparison between two-dimensional and three-dimensional cephalometry on frontal radiographs and on cone beam computed tomography scans of human skulls. *Eur. J. Oral Sci.* **2009**, *117*, 300–305. [CrossRef] [PubMed]

biology

MDPI

Article

Forensic Tools for Species Identification of Skeletal Remains: Metrics, Statistics, and OsteoID

Heather M. Garvin [1,*], Rachel Dunn [1], Sabrina B. Sholts [2], M. Schuyler Litten [2], Merna Mohamed [3], Nathan Kuttickat [3] and Noah Skantz [3]

[1] Department of Anatomy, Des Moines University, Des Moines, IA 50312, USA; rachel.dunn@dmu.edu
[2] National Museum of Natural History, Smithsonian Institution, Washington, DC 20056, USA; SholtsS@si.edu (S.B.S.); littenm@si.edu (M.S.L.)
[3] College of Osteopathic Medicine, Des Moines University, Des Moines, IA 50312, USA; merna.mohamed@dmu.edu (M.M.); nathan.kuttickat@dmu.edu (N.K.); noah.a.skantz@dmu.edu (N.S.)
* Correspondence: heather.garvin-elling@dmu.edu

Simple Summary: Forensic anthropologists are commonly asked to determine whether bones are of human origin and, if not, to which species they belong. Current practice usually relies on visual assessments rather than quantitative analyses. This study aimed to test the utility of basic bone metrics in discriminating human from nonhuman elements and assigning faunal species. A database of more than 50,000 skeletal measurements was compiled from humans and 27 nonhuman species. Equations and classification trees were developed that can differentiate human from nonhuman species with upwards of 90% accuracy, even when the bone type is not first identified. Classification trees return accuracy rates greater than 98% for the human sample. These quantitative models provide statistical support to visual assessments and can be used for preliminary assessment of a bone's forensic significance at a scene. The statistical models, however, could not classify species at acceptable rates. For species identification, a freely available web tool (OsteoID) was created from the study data, where users can filter photographs of potential bones/species using a few basic measurements and access 3D scans and additional resources to facilitate identification. OsteoID provides an important resource for forensic anthropologists lacking access to large comparative skeletal collections, as well as other disciplines where comparative osteological training is necessary.

Abstract: Although nonhuman remains constitute a significant portion of forensic anthropological casework, the potential use of bone metrics to assess the human origin and to classify species of skeletal remains has not been thoroughly investigated. This study aimed to assess the utility of quantitative methods in distinguishing human from nonhuman remains and present additional resources for species identification. Over 50,000 measurements were compiled from humans and 27 nonhuman (mostly North American) species. Decision trees developed from the long bone data can differentiate human from nonhuman remains with over 90% accuracy (>98% accuracy for the human sample), even if all long bones are pooled. Stepwise discriminant function results were slightly lower (>87.4% overall accuracy). The quantitative models can be used to support visual identifications or preliminarily assess forensic significance at scenes. For species classification, bone-specific discriminant functions returned accuracies between 77.7% and 89.1%, but classification results varied highly across species. From the study data, we developed a web tool, OsteoID, for users who can input measurements and be shown photographs of potential bones/species to aid in visual identification. OsteoID also includes supplementary images (e.g., 3D scans), creating an additional resource for forensic anthropologists and others involved in skeletal species identification and comparative osteology.

Keywords: forensic anthropology; medicolegal death investigation; forensic significance; comparative osteology; human osteology; skeletal morphology; nonhuman

Citation: Garvin, H.M.; Dunn, R.; Sholts, S.B.; Litten, M.S.; Mohamed, M.; Kuttickat, N.; Skantz, N. Forensic Tools for Species Identification of Skeletal Remains: Metrics, Statistics, and OsteoID. *Biology* **2022**, *11*, 25. https://doi.org/10.3390/biology11010025

Academic Editors: Ann H. Ross, Eugénia Cunha and Andrés Moya

Received: 31 October 2021
Accepted: 22 December 2021
Published: 25 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Forensic anthropologists are commonly approached by law enforcement, coroners, and medical examiners with an unknown skeletal element and faced with a simple question: is this human [1,2] Well-trained forensic anthropologists know the human skeletal system in meticulous detail, and unless the skeletal element has been highly modified (e.g., extreme fragmentation, burning, etc.), they can usually differentiate human from nonhuman remains without hesitation [3]. Forensic anthropologists visually assess the bone, determining the element type (e.g., humerus, femur, tibia, etc.) and whether it is consistent with human anatomy based on its size (given its developmental state), shape, and bony features [3]. This macroscopic assessment is usually concluded without metric analyses.

If the bone is human, it is of forensic significance and will be subjected to a comprehensive osteological analysis. If the bone is nonhuman, a forensic anthropologist is faced with an inevitable follow-up question: what is it? This question is more than mere curiosity because it provides verifiable evidence to support the forensic anthropologist's nonhuman designation [3]. An incorrect faunal species identification can affect the forensic anthropologist's credibility, even if it is not of forensic importance. Similarly, responding to the inquiry by stating that it is not important or that you do not know does not instill confidence or foster positive relationships with agencies. In some cases, the animal species may provide investigators additional evidence or context regarding the circumstances of death. For example, if the remains of a cat are found intermixed with human remains, it may suggest that a suspect disposed of a house pet along with the decedent in an attempt to conceal the human remains.

Faunal species identification, however, can be challenging for practitioners given the number of bones in a skeleton, variety of potential species, and similar morphology amongst related species [4]. While forensic anthropologists are required to be experts on the human skeleton, zooarchaeological training, while ideal, is not a requirement, and expertise in comparative osteology can vary greatly amongst practitioners. When determining the nonhuman species of skeletal remains, practitioners are fortunate if they have access to comparative osteological collections to assist with identifications. Such collections take time and resources to build or require proximity and unrestricted accessibility to an already-established collection. Various comparative osteology texts are available [5–13], each with their own advantages and limitations; they vary in cost, comprehensiveness, species included, photographic quality, and target audience. Texts are also most useful if the user knows the element type in advance and/or already suspects a certain species. Reliable and easily accessible online resources are limited, and internet searches for images of specific faunal elements can return mixed results.

The primary goal of this project was to develop additional, freely-available resources to support forensic anthropologists and medicolegal personnel in skeletal species identification based on simple measurements. Saulsman et al. [14] report discriminant functions derived from eight traditional long bone metrics that can differentiate human from five Australian nonhuman species with accuracy rates at or above 95%. Their sample sizes were limited to 50 human and 50 nonhuman individuals (ten per species). Given their promising results, this study aimed to test the utility of similar bone metrics in differentiating much larger samples of human and nonhuman specimens and classifying species, with a focus on species commonly encountered in North America. Although a handful of measurements cannot capture specific distinguishing bony features, traditional morphometric analyses can capture overall bone size and shape (i.e., form), which are variables considered subjectively during visual assessments of species.

In addition to the morphometric analyses, this study also aimed to develop a freely available searchable online database that uses basic metrics and visual aids (i.e., photographs and 3D scans) to help forensic anthropologists and medicolegal personnel (amongst others) determine species from skeletal elements. These resources would benefit practitioners without access to extensive comparative collections and would be accessible in the field via the use of a smart phone or other device. Beyond the scope of forensic anthropology,

this skeletal species identification tool may be useful to students, archaeologists, wildlife forensic specialists, biologists, veterinarians, and others, including the general public who may wish to learn more about bones they encounter through various activities.

2. Materials and Methods

The study sample included skeletal data from humans and 27 faunal species frequently found in North America (20 mammals, 5 birds, 2 turtles—see Table 1), which included species that approximate human sizes (e.g., deer, horse, elk, moose, cow, pig, domestic dog, and black and brown bears). The species included are also commonly presented in comparative osteology texts used by forensic anthropologists [5–9] and encountered in forensic anthropological analyses [1]. To facilitate database searching, analogous measurements needed to be obtainable from each specimen included, regardless of species or element type. Thus, long bones were chosen as the main focus for this study (humerus, radius, ulna, radio-ulna, femur, tibia, fibula, and fused metapodials). For birds, the tibiotarsus was included with the tibia data, and the carpometacarpus and tarsometatarsus were included with the fused metapodials. The scapula, sacrum and os coxae were also included given the ability to take maximum lengths and breadths and their diagnostic morphologies. The original measurement list consisted of maximum lengths, proximal and distal maximum breadths (medio-lateral) and depths (antero-posterior), midshaft minimum and maximum diameters, and a few unique measurements for certain elements (e.g., femoral head diameter, acetabular diameter). Von den Driesch [15] was used as a guide when establishing the measurements.

These measurement data were collected from skeletal remains curated at the following institutions: Smithsonian National Museum of Natural History, Washington, DC; American Museum of Natural History, New York City, NY; Mercyhurst University, Erie, PA; Washburn University, Topeka, KS; University of California, Davis, CA; and Des Moines University, Des Moines, IA. Additional data were included from published papers and available datasets [16–34]. In some cases, published data of specimens outside of North America were included in the study to increase sample sizes if the species was the same as that commonly encountered in North America (e.g., domestic dogs and cats). Inclusion in the study required specimens to be of skeletal maturity; specimens in advanced stages of epiphyseal fusion were included to increase faunal sample sizes where necessary. This original dataset consisted of 59,442 measurements from 18,867 bones from 5207 individuals/animals). Species averages, standard deviations, and minimum/maximum ranges were calculated for each measurement. Photographs of exemplar specimens were taken from multiple standard views (e.g., six views for long bones) for incorporation into the web tool.

A subset of the data (47,688 measurements collected from 16,315 long bone elements) was subjected to linear discriminant function (DFA) and decision tree analyses to evaluate potential methods of human versus nonhuman and species classifications (Table 1). This subset included maximum length (MaxL), maximum mediolateral width of the proximal epiphysis (MaxPW), maximum mediolateral width of the distal epiphysis (MaxDW), maximum anteroposterior depth of the distal epiphysis (MaxDD), maximum diameter of the midshaft (MaxMidD), and minimum diameter of the midshaft (MinMidD) collected from humeri, radii, ulnae, femora, and tibiae. Element-specific measurements (e.g., femoral head diameter) were excluded to permit pooled analyses across element types. Maximum proximal depth was excluded due to measurement difficulty in certain elements (e.g., tibia depending on tuberosity location, ulna, and radio-ulna). Step-wise DFA using Wilk's lambda and a leave-one-out cross-validation were performed on the human versus pooled nonhuman samples of all long bones (replicating a situation where the element type is unknown), and then separately for each bone. DFA was used to assess human versus nonhuman classification for commonly collected univariate variables (MaxL, MaxPW, and MaxDW) and variables grouped by bone region (e.g., distal measurements and midshaft measurements) for application in cases when the unknown element is incomplete/fragmented or taphonomic modifications preclude some measurements. Finally,

stepwise discriminant functions were also run to assess potential ability to classify the 28 species using both pooled-bone and bone-specific samples. Variables input into the stepwise analyses were chosen to maximize sample sizes and discriminatory power. Box's M was used to assess homogeneity in variance–covariance matrices, and Kolmogrov–Smirnov tests were performed to evaluate data normality.

Table 1. List of species from which data and photos were collected and sample sizes by element.

Class	Genus	Species	Common Name	Humerus	Femur	Radius	Tibia	Ulna [1]
Aves	Anas	platyrhynchos	Mallard Duck	31	28	31	30	31
Aves	Aquila	chrysaetos	Golden Eagle	21	23	20	19	23
Aves	Branta	canadensis	Goose	34	34	32	31	34
Aves	Gallus	gallus	Chicken	31	31	31	32	31
Aves	Meleagris	gallopavo	Turkey	35	35	32	35	34
Mammalia	Alces	alces	Moose	19	17	20	21	27
Mammalia	Bos	taurus	Cow	15	16	13	17	12
Mammalia	Canis	familiaris	Domestic Dog	84	147	75	147	76
Mammalia	Canis	latrans	Coyote	64	65	57	65	58
Mammalia	Canis	lupus	Wolf	44	45	38	45	38
Mammalia	Capra	hircus	Goat	83	3	79	80	3
Mammalia	Cervus	canadensis	Elk	34	33	31	32	31
Mammalia	Didelphis	virginiana	Opossum	34	34	35	33	33
Mammalia	Ovis/capra [2]	aries/hircus	Sheep/Goat	2	1	1	0	1
Mammalia	Equus	caballus	Horse	31	33	33	30	33
Mammalia	Felis	catus	Domestic Cat	40	39	39	39	38
Mammalia	Homo	sapiens	Human	2714	2700	2672	2684	463
Mammalia	Odocoileus	hemionus	Mule deer	31	32	34	32	38
Mammalia	Odocoileus	virginianus	White-Tailed Deer	33	39	35	39	35
Mammalia	Ovis	aries	Sheep	77	18	147	104	63
Mammalia	Procyon	lotor	Racoon	36	37	36	39	37
Mammalia	Sus	scrofa	Domestic Pig/Boar	20	17	7	17	8
Mammalia	Sylviagus	floridanus	Eastern Cotton-Tail Rabbit	36	34	34	32	33
Mammalia	Urocyon	cinereoargenteus	Gray Fox	39	42	39	40	42
Mammalia	Ursus	americanus	American Black Bear	38	34	18	19	18
Mammalia	Ursus	arctos	Brown Bear	48	46	18	22	19
Mammalia	Vulpes	vulpes	Red Fox	43	41	41	42	40
Testudines	Chelydra	serpentina	Snapping Turtle	30	30	30	30	30
Testudines	Terrapene	carolina	Common Box Turtle	31	31	27	31	31
			Totals	3778	3685	3705	3787	1360

[1] For the human and nonhuman comparisons, individual measurements were taken from fused radio-ulna elements and included as radius or ulna. For the development of the web tool, both the individual radius and ulna measurements and combined maximum lengths/widths for the fused radio-ulna were included for search purposes. [2] A few specimens were labeled as "Sheep/Goat" in the collection and thus entered this way for human versus nonhuman analyses but were excluded from species analyses.

Decision trees were developed from the same data set and evaluated for classifying human versus the pooled nonhuman samples and classifying species using both the pooled-bone sample and bone-specific subsamples. The decision trees were created using a CRT (Classification and Regression Trees) growth model with a Gini impurity measure splitting criterion and a maximum tree depth of five levels. CRT uses stepwise variable selection to create a decision tree where each node is split using the variable that best maximizes the purity of the resulting nodes (i.e., homogeneity of the dependent variable) [35,36]. CRT also uses surrogate variables (those that result in a similar outcome pattern) to replace missing data, thereby maximizing sample sizes. The minimum number of cases for nodes was set at 100 for parent nodes and 50 for child nodes. Equal prior probabilities were used across groups. Tree pruning was implemented, set at one standard error in order to avoid overfitting [35,36]. A split-sample validation was applied, with the model generated from a training sample (70% of the data), which was then validated on the test sample (remaining 30% of the data). For the trees classifying human from nonhuman remains, human was set

as a target variable and a misclassification cost of ten was assigned to misclassifications of human bone as nonhuman. This reflects the more severe forensic implications in erroneously assigning a human bone as nonhuman as compared to misclassifying a nonhuman bone as human.

The linear discriminant function analyses represent more traditional classification approaches but have statistical assumptions such as multivariate normality and homogeneity of variance–covariance matrices [37–39]. Decision trees do not rely on these statistical assumptions [40–42]. All statistical analyses were performed in SPSS v.28 (IBM Corporation, Armonk, NY, USA). We hypothesized that the multivariate DFA and decision trees would be able to adequately differentiate human from nonhuman remains when single elements were assessed, given that these morphometric parameters are used during visual assessments of remains. The pooled-bone sample is expected to provide less accurate results, given the compounded effects of variation within and between species and element types. The results of the DFA and decision trees were used to make informed decisions about the development of the skeletal species identification web tool, with the possibility of integrating the methods into the tool depending on their performance.

3. Results

3.1. Descriptive Statistics

Sample sizes, minimum and maximum values, averages, standard deviations, and the ranges between two negative and two positive standard deviations (~95% confidence interval) were calculated per measurement and species (38 measurements collected across 28 species). Given the forensic aim to distinguish human from nonhuman remains, as well as the extensive dataset, Table 2 presents only the human summary statistics. This table may act as a general guide to assess whether a bone falls within the human size ranges; note, however, that there is always a small possibility of a human bone falling outside these values, given that samples may not represent the complete global variation of past and present populations. Descriptive statistics for nonhuman measurements by species are provided in the Supplementary Materials (Tables S1–S11).

3.2. Morphometric Human Versus Nonhuman Classification

When the human long bone measurements are compared to those of the pooled nonhuman long bones, Box's M indicates significant differences in the variance–covariance matrices ($p < 0.001$ for all analyses). This is true for both the pooled-bone and bone-specific samples. Kolmogrov–Smirnov results indicate that the nonhuman variables are not normally distributed, while the human data generally do not differ significantly from normality ($p > 0.05$). These results are unsurprising given the unequal sample sizes and range of nonhuman species being pooled (Table 1). DFA has been suggested to be robust against statistical violations [42]. For this reason and the exploratory nature of the analyses, the DFAs were performed despite the violation of statistical assumptions to provide comparison to the decision tree results and informed decisions about the web tool development.

The results of the human versus nonhuman DFA classification are summarized in Table 3, including overall cross-validated accuracy, group-specific cross-validated correct classifications, and sample sizes for each model. Note that DFA requires that all measurements are present for each element in the analysis, resulting in significant decreases in sample sizes for some models due to missing data. In each analysis, the cross-validated results were the same or similar to the original classification results. There are some classification biases, but in most cases, the human correct classification is higher than the nonhuman. Of the univariate analyses, maximum lengths performed the best with overall classification rates above 90% for all elements except for the ulna and a 79.5% classification rate for the pooled-bone analysis. The human classification rates using only maximum length were over 99% for all bones except the ulna (96.8%). The DFAs assessing regional measurements (two midshaft variables or two distal variables) provided results similar to or lower than the univariate maximum length results, with a

few exceptions. The ulna midshaft had a 90.0% correct classification, outperforming the length results, and the humerus midshaft accuracy was much lower than the length at 67.1% (vs. 94.1% for maximum length).

Table 2. Descriptive statistics from the human sample, including counts, minimums, maximums, averages, standard deviations, and two standard deviation ranges per element and measurement.

Bone	Meas [1]	N	Min	Max	Ave	SD	−2SD	+2SD
Humerus	MaxL	2567	225	397	309	23	262	356
	MaxPD	94	35	56	46	4	37	55
	MaxPW	411	38	62	49	4	41	58
	MaxDD	94	22	37	28	3	22	35
	MaxDW	1867	42	77	59	6	47	70
	MidMaxD	425	16	32	23	3	17	28
	MidMinD	440	11	24	18	2	13	23
Radius	MaxL	2531	180	309	236	20	196	276
	MaxPD	380	15	31	23	2	18	28
	MaxPW	380	15	31	23	2	18	27
	MaxDD	89	17	36	25	4	17	32
	MaxDW	89	21	42	33	4	25	41
	MidMaxD	454	10	21	16	2	12	19
	MidMinD	454	8	74	12	3	5	19
Ulna	MaxL	406	211	334	263	20	222	303
	MaxPW	257	14	35	26	3	19	32
	MidMaxD	477	10	24	17	2	12	21
	MidMinD	477	9	19	13	2	9	17
Femur	MaxL	2630	344	550	433	33	367	499
	MaxPD	89	37	59	46	5	36	56
	MaxPW	89	71	105	87	8	72	103
	DiamH	1077	35	61	44	4	37	52
	MaxDD	89	46	92	63	7	49	78
	MaxDW	2563	58	98	77	6	64	90
	MidMaxD	457	14	39	27	3	21	33
	MidMinD	457	17	39	27	3	21	33
Tibia	MaxL	2589	159	472	357	32	294	421
	MaxPW	1867	50	94	71	6	58	84
	MaxDD	82	30	52	39	4	31	47
	MaxDW	415	40	63	52	4	43	60
	MidMaxD	420	19	44	33	4	25	42
	MidMinD	82	15	28	21	3	16	26
Fibula	MaxL	429	282	463	366	27	312	421
Os Coxae	MaxL	91	166	237	202	16	170	233
	DiamA	1526	39	63	49	4	41	57
Sacrum	MaxL	90	89	157	114	13	88	141
	MaxPW	90	90	138	111	9	93	129
Scapula	MaxL	92	127	210	178	17	145	212

[1] Measurement abbreviations: MaxL = maximum length, MaxPW = maximum proximal width (medio-lateral), MaxDD = maximum distal depth (antero-posterior), MaxDW = maximum distal width (medio-lateral), MidMaxD = maximum diameter at midshaft, MidMinD = minimum diameter at midshaft, DiamH = femoral head diameter, DiamA = acetabulum diameter.

As expected, the pooled-bone DFAs did not perform as well as the bone-specific analyses for morphometric human versus nonhuman classification. The pooled-bone univariate analysis of maximum distal width performed the best (87.9%), which may be because ulnae were excluded from this analysis (distal ulna measurements were not collected) thereby removing one confounding element. Maximum length correctly classified 79.5% of the sample composed of 11,129 human bones and 5254 nonhuman bones.

Table 3. Linear DFA accuracy results and sample sizes for human (Hum) and nonhuman (Non) classifications summarized by element and variables. Overall accuracy is bolded. Var(s) = variable(s), N_{Hum} = human sample size, N_{Non} = nonhuman sample size.

Var(s)	Pooled-Bone	Humerus	Femur	Radius	Tibia	Ulna
MaxL	**79.5%** Hum: 79.3%; Non: 80.0% N_{Hum} = 11,129; N_{Non} = 5254	**94.1%** Hum: 99.9%; Non: 78.0% N_{Hum} = 2567; N_{Non} = 920	**95.4%** Hum: 100.0%; Non: 83.2% N_{Hum} = 2630; N_{Non} = 981	**95.4%** Hum: 99.8%; Non: 79.3% N_{Hum} = 2531; N_{Non} = 697	**95.0%** Hum: 99.5%; Non: 83.9% N_{Hum} = 2589; N_{Non} = 1062	**78.5%** Hum: 96.8%; Non: 69.1% N_{Hum} = 406; N_{Non} = 797
MaxPW	**74.8%** Hum: 68.8%; Non: 79.2% N_{Hum} = 3383; N_{Non} = 4668	**75.4%** Hum: 85.4% Non: 69.8% N_{Hum} = 410; N_{Non} = 733	**85.4%** Hum: 100.0%; Non: 83.8% N_{Hum} = 89; N_{Non} = 822	**46.1%** Hum: 61.8%; Non: 39.7% N_{Hum} = 380; N_{Non} = 946	**93.4%** Hum: 99.7%; Non: 78.8% N_{Hum}: 1867; N_{Non} = 817	**74.3%** Hum: 70.8%; Non: 76.5% N_{Hum} = 257; N_{Non} = 404
MaxDW	**87.9%** Hum: 92.9%; Non: 82.1% N_{Hum} = 5021; N_{Non} = 4345	**92.0%** Hum: 97.9%; Non: 81.2% N_{Hum} = 1868; N_{Non} = 1024	**94.8%** Hum: 100.0%; Non: 80.3% N_{Hum} = 2560; N_{Non} = 908	**69.7%** Hum: 83.1% Non: 68.1% N_{Hum} = 89; N_{Non} = 745	**89.6%** Hum: 100.0%; Non: 84.9% N_{Hum} = 415; N_{Non} = 923	–
MaxDD & MaxDW	**78.9%** Hum: 64.3%; Non: 80.0% N_{Hum} = 230; NHn = 3007	**96.3%** Hum: 100.0%; Non: 96.21% N_{Hum} = 26; N_{Non} = 850	**96.1%** Hum: 97.0%; Non: 96.1% N_{Hum} = 33; N_{Non} = 720	**86.4%** Hum: 82.0%; Non: 87.3% N_{Hum} = 89; N_{Non} = 448	**86.2%** Hum: 92.7% Non: 85.4% N_{Hum} = 82; N_{Non} = 735	–
MidMaxD & MidMinD	**64.4%** Hum: 49.0%; Non: 73.1% N_{Hum} = 1767; N_{Non} = 3089	**67.1%** Hum: 62.0%; Non: 69.9% N_{Hum} = 347; N_{Non} = 714	**90.2%** Hum: 86.4%; Non: 92.7% N_{Hum} = 457; N_{Non} = 711	**87.9%** Hum: 88.3%; Non: 87.6% N_{Hum} = 436; N_{Non} = 443	**87.2%** Hum: 84.8% Non: 87.5% N_{Hum} = 66; N_{Non} = 537	**90.0%** Hum: 94.4%; Non: 85.3% N_{Hum} = 461; N_{Non} = 428
Stepwise [1]	**90.3%** Hum: 95.6%; Non: 87.9% N_{Hum} = 1408; N_{Non} = 3088 MaxL, MidMaxD, MidMinD	**96.7%** Hum: 99.6%; Non: 90.7% N_{Hum} = 1862; N_{Non} = 891 MaxL, MaxDW	**98.1%** Hum: 99.9%; Non: 93.0% N_{Hum} = 2552 N_{Non} = 906 MaxL, MaxDW	**91.4%** Hum: 100.0%; Non: 86.8% N_{Hum} = 327; N_{Non} = 621 MaxL, MaxPW	**89.4%** Hum: 92.2%; Non: 83.4% N_{Hum} = 1773 N_{Non} = 807 MaxL, MaxPW	**87.4%** Hum: 93.7%; Non: 77.7% N_{Hum} = 254; N_{Non} = 166 MaxL, MaxPW, MidMaxD, MidMinD

[1] All variables were included in the stepwise DFA and those retained in the function are listed in each column with the results.

The multivariate stepwise DFAs returned correct human versus nonhuman classification rates above 90% for the humerus, femur, and radius and just below 90% for the tibia and ulna (Table 3). Maximum length was utilized in all the stepwise functions and had the highest weight. For the humerus (n = 2753) and femur (n = 3458), a function including maximum length and maximum distal width returned accuracy rates of 96.7% and 98.1%, respectively. Other functions for the humerus and femur returned higher classification rates (99.5% for the humerus and 99.7% for the femur), but given the variables included in these functions, sample sizes decreased to around 1100. Equations associated with the multivariate discriminate functions are provided in the Supplementary Materials (Table S12).

The decision tree results outperformed the DFA results for human versus nonhuman classification (Table 4) and were derived from larger samples in both the training and test sets. With all bones pooled, decision trees that evaluated all measurements correctly classified 90% or more of the training and test samples, except for the ulna test sample (89.3%). The region-specific pooled-bone analyses had lower accuracy rates (ranging from 76 to 89% correct) but still outperformed the DFA. With the exception of the ulna test sample, all training and test samples had correct human classification rates of 98% or higher.

The ulna test sample correctly classified 94.5% of the human sample. Using four basic measurements, the decision tree presented in Figure 1 results in an overall accuracy of 91% and human classification accuracy of 99.6%; this is for the pooled-bone sample (i.e., without first identifying which bone is present). Although the nonhuman classification rate is lower (75%), this bias is expected given that we assigned higher misclassification costs to the human sample. The terminal nodes of the decision tree (Figure 1) indicate the number/percentage of human and nonhuman elements that fell within that node as well as associated sample sizes. Note that the "total" row depicts the percentage of the original input sample. The terminal nodes vary in their accuracy rates (75.2 to 99.8%), but only one of five terminal nodes had accuracy rates below 90%. This node (node 7) consists of ~17% of the total sample and represents those elements in which the multivariate sizes overlap between human and nonhuman species. For example, a deer metatarsal may approximate a human radius based on the measurements. Decision trees associated with the results in Table 4 are presented in the Supplementary Materials (Figures S1–S9).

Table 4. Decision tree results and sample sizes for human (Hum) and nonhuman (Non) classifications summarized by element and variables. Acc = accuracy, N = sample size. See Supplementary Materials (Figures S1–S9) for the decision trees.

Bone	Input	Training Sample					Test Sample					Tree Variables
		Total Acc	Hum Acc	Non Acc	Hum N	Non N	Total Acc	Hum Acc	Non Acc	Hum N	Non N	
Pooled	All 6	91.4%	99.6%	75.4%	8211	4253	91.0%	99.6%	75.0%	3487	1865	MaxL, MidMaxD, MaxDW & MaxPW
Pooled	Distal	83.3%	98.8%	64.8%	3650	3052	82.9%	98.9%	64.5%	1495	1300	MaxDD & MaxDW
Pooled	Mid	77.2%	99.7%	61.9%	1603	2356	75.7%	99.4%	59.6%	689	1016	MidMaxD & MidMinD
Pooled	Length	88.3%	99.9%	64.5%	7696	3763	88.7%	99.8%	63.2%	3433	1491	MaxL
Humerus	ALL 6	99.1%	99.3%	98.4%	1914	741	97.9%	98.6%	96.3%	776	323	MaxL & MidMinD
Femur	ALL 6	96.7%	99.4%	89.5%	1853	694	96.4%	99.6%	86.9%	837	290	MaxL only
Radius	ALL 6	95.1%	98.6%	86.2%	1894	734	94.9%	98.6%	85.2%	778	298	MaxL & MidMaxD
Tibia	ALL 6	94.9%	98.4%	86.5%	1854	776	94.0%	97.8%	84.4%	830	327	MaxDW & MaxPW
Ulna	ALL 4	92.5%	98.7%	89.2%	318	590	89.3%	94.5%	86.5%	145	267	MidMinD & MaxPW

3.3. Morphometric Skeletal Species Identification

Correct species classification rates from the stepwise DFAs are summarized in Table 5. The pooled-bone analysis had an overall 40.4% accuracy rate, which, although better than the a priori classification rate (3.6%), can lead to numerus classification issues. For this model, 20 species had correct classification rates below 50%, with only two species (eastern cotton-tail rabbit and common box turtle) with classification rates above 75% (both above 90%). Bone-specific DFAs performed better, with overall accuracies ranging from 78 to 89%. The humerus DFA had the most accurate classifications with 18 species above 90% and none below 50%. The humerus DFA performed the worst for brown bear (55.6%), domestic dog (53.7%), and pig (50.0%). Domestic dog had classification issues across all DFAs given the high degree of variation in dog sizes and morphologies. Species within the same genus were commonly misclassified (e.g., domestic dogs and coyotes, brown bears and black bears, etc.), given their similarity in morphology and substantial overlap in body size. Human classification rates for the bone-specific DFAs ranged from 76.8% (ulna) to 100.0% (humerus, femur, and radius). All stepwise DFAs retained all variables in the final functions, and maximum length was consistently the most important variable. Ultimately, while the overall species classification rates for the bone-specific DFAs are acceptable, results varied greatly by taxa, suggesting that the DFAs should only be used as a general guide and should not be relied on as final determinants of species identification.

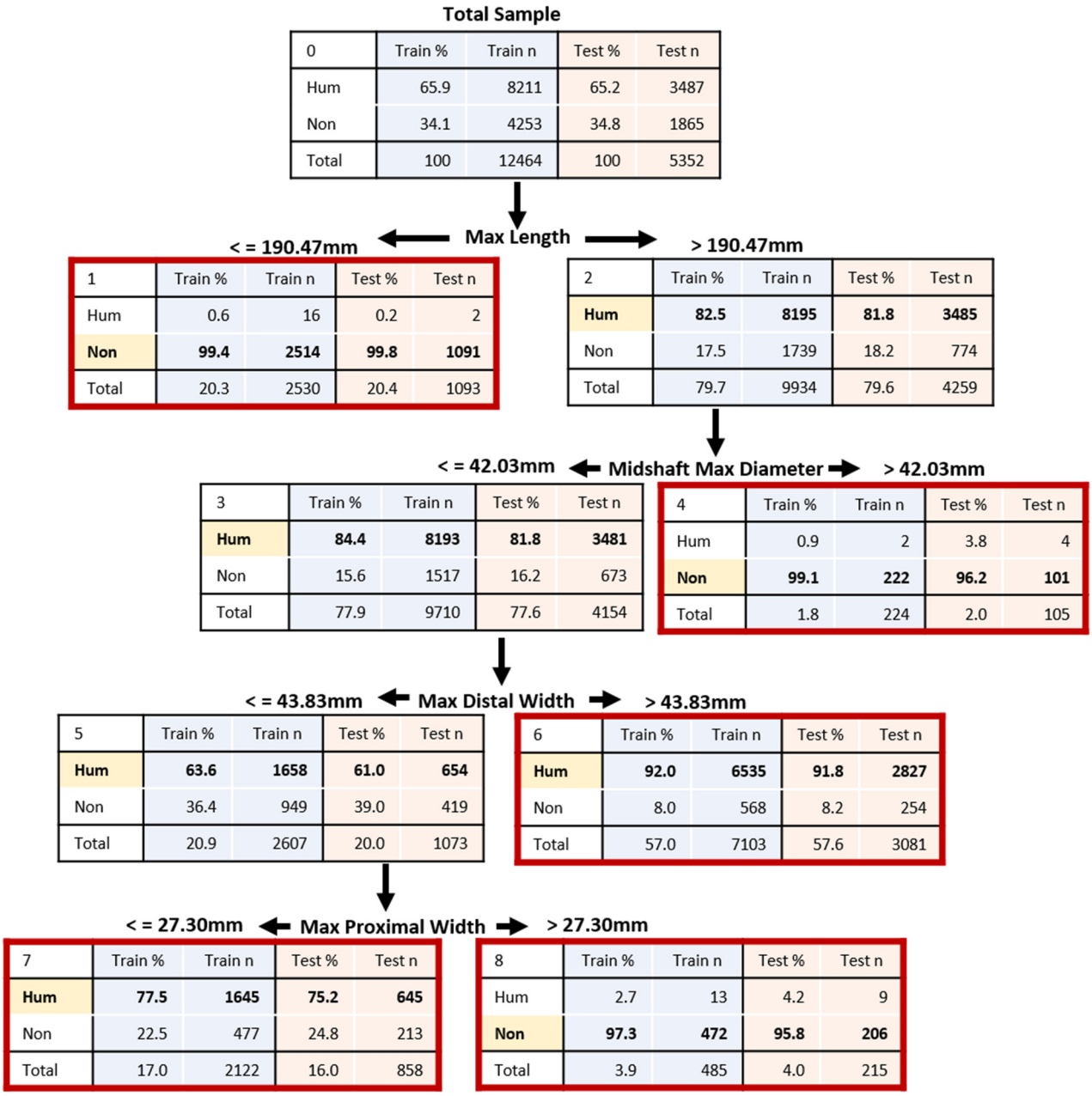


Figure 1. Decision tree developed to classify human (Hum) versus nonhuman (Non) elements from a pooled-bone sample (i.e., all long bones pooled). Working from the top of the tree, the variable listed at each level would be measured, and based on the provided sectioning point, the user would move down the tree to the next level. This process would continue until arriving at a terminal node where classification would be assigned. Terminal nodes are outlined in red. Group classification is highlighted in yellow and bolded at each node. Percentages and counts of bones classified to each group in the training and testing samples are presented, as well as the total percentage of the sample represented in that node. Overall correct classification for the test sample is 91.0% (99.6% for human and 75.0% for nonhuman elements). This decision tree corresponds with the first line in Table 4.

As might be expected, the decision tree results attempting to classify species were not successful. While tree overall classification rates were over 70% for all analyses except the ulna, none of the trees produced 28 terminal nodes to classify each species. To classify each species would require too many levels and branches; thus, the trees opted for preserving overall classification rates by focusing on those species with the highest counts.

Table 5. Stepwise DFA species classification results. The right side of the table presents the number of species that fell within each accuracy range (i.e., <50%, 50–75%, 75–90%, or >90%). Vars = variables included in final function, Acc = Accuracy, Hum = human.

Bone	N	Vars	Total Acc	Hum Acc	Number of Species			
					<50%	**50–75%**	**75–90%**	**>90%**
Pooled	2737	All 6	40.4%	68.9%	20	6	0	2
Humerus	735	All 6	89.1%	100.0%	0	6	4	18
Femur	744	All 6	79.3%	100.0%	1	4	14	8
Radius	462	All 6	83.9%	100.0%	4	5	8	10
Tibia	548	All 6	77.7%	92.9%	3	5	5	14
Ulna	420	All 4	79.2%	76.8%	4	2	6	7

3.4. Web Tool for Species Identification

Both the DFA and decision tree results suggest that a simple equation or tree cannot be used to adequately identify skeletal species. When forensic anthropologists visually evaluate skeletal remains, they mentally process the bone dimensions to consider possible species, using the overall bone size and shape to narrow down potential species. Ultimately, however, visual comparisons and specific bony features are used to make final species identifications.

To facilitate this species identification process, we utilized the metric data and images from our study sample to develop an online, freely available species identification tool: OsteoID [43]. The home page asks users to first identify the bone, providing diverse exemplars for each element (humerus, femur, radius, radio-ulna, ulna, tibia, fibula, metapodials, scapula, sacrum, and os coxae), demonstrating the common general morphology of specific elements across most species. There is also an option to "Search All" if the user cannot confidently determine bone type. Once an option is selected, the user is brought to a new page where they can narrow the search by common name, scientific name, or by bone length, proximal width, and distal width. At any point, the user can search additional fields in the side bar.

Maximum length, proximal width and distal width were chosen as the web tool filtering variables for several reasons. First, they were found to be the easiest to measure reliably, even with little or no osteological experience. In addition, the DFA and decision tree analyses revealed maximum length to be the most important variable in species identification, followed commonly by maximum distal width; including distal depth did not exclude many more species. Finally, the midshaft measurements are instrumentally defined (i.e., users need to take the maximum length and divide it by two to determine the correct location to take the midshaft maximum and minimum diameters) and require calipers. These factors make application in the field difficult and limit utility to those with osteological backgrounds.

To determine the searchable range for each species/bone measurement, the minimum, maximum, and two standard deviations above and below the mean were calculated. The smallest value (whether two standard deviations below the mean or the observed minimum) was used as the lower search limit, while the largest value (either two standard deviations above the mean or the observed maximum) was used as the upper search limit. This created a conservative size range, which is important given that the dataset does not likely encompass the full size range of each species. For elements in the database missing one or more measurements, a range of 0–1000 mm was assigned so that it would not be automatically eliminated during searches.

As possible bones/species are narrowed, thumbnails show multi-views of the bones by species as well as a list of the possible measurement ranges. Clicking on the thumbnails opens a larger image in a new window. By opening in a new window, multiple possible matches can be opened and placed side-by-side if needed. Most figures have six views of the exemplar element (anterior, posterior, medial, lateral, proximal, and distal) and include

the maximum length range on the image, a scale bar, and, when possible, a penny was added for more intuitive sizing. Genus, species, collection, bone, and side information is also provided. Some images have been annotated to point out distinctive features. The user ultimately makes their final species classification based on visual comparisons. This web tool is also compatible for use on smartphones and thus is accessible in the field.

Informational tabs on the home screen describe the web tool and its development, provide instructions on utilizing the web tool (including measurement images), and answer frequently asked questions. Users are reminded that filtering the bones/species by measurements only works for skeletally mature specimens and are instructed on how to identify skeletal maturity. In numerus places, users are reminded that if a bone has any possibility of being human, they need to contact the local law enforcement agency immediately.

Finally, a tab also refers the user to additional resources [43]. This includes references to other texts or websites as well as a link to a Dropbox folder where they can find additional project resources. In this folder, users can find the images included in the web tool, as well as images of other elements such as carpals and tarsals, which were not included in the main web tool given that measurements were not collected from these elements. Three-dimensional surface scans of many of the elements are also provided, which can be downloaded by users to view for comparison or 3D print. These 3D prints may be used to build or supplement comparative osteology collections. We are continuously expanding these Supplementary Materials and uploading them to additional digital repositories (e.g., [44]). Finally, the project data can also be accessed in this folder, as well as on Dryad [45].

4. Discussion

4.1. Human Versus Nonhuman Determination

Nonhuman remains comprise a significant portion (25–30%) of total cases assessed by forensic anthropologists [1–3] and can represent more than 90% of skeletal cases submitted to medical examiner offices [1]. Although forensic anthropologists mentally assess bone size and shape when determining skeletal species, only one other published study was found that assessed the utility of basic long bone osteometrics in differentiating human from nonhuman remains. Saulsman et al. [14] created discriminate functions from a sample of 50 human and 50 nonhuman specimens from five Australian species. Their study illustrated the potential utility of such quantitative methods, with accuracy rates over 95%, but it was limited by sample sizes and species inclusion.

Our results, where more than 16,000 long bones were assessed quantitatively to develop predictive models, support their findings. From this extensive dataset, we provide discriminant functions and decision trees that can be used to assist or support human versus nonhuman determinations from long bones. Even when all elements are pooled, the DFA and decision trees return over 90% accuracy, with correct classifications of human remains over 95% (99.6% for the decision tree). Thus, high accuracy rates can be achieved even without first distinguishing the specific bony element present. If the bone is first identified and bone-specific methods are applied, accuracy increases further for all models except the tibia-specific and ulna-specific discriminant functions, which were slightly lower. The ulna performed the worst across most analyses, which may partly be due to the lack of distal measurements collected for this element. Generally, the decision tree presented slightly higher overall accuracy rates as compared to the DFAs.

When assessing the human versus nonhuman origin of skeletal remains, we recommend the use of the decision trees presented in this paper and Supplementary Materials compared to the discriminant functions, given (1) their higher accuracy rates, (2) their use of more available data and split-validation, and (3) their lack of statistical assumptions [42]. The better performance of decision trees may also reflect the incorporation of multiple sectioning points into the model (one at each node) as compared to a single sectioning point with discriminant functions. In addition, decision trees provide classification rates

at each of the nodes, providing a more realistic view of accuracy and confidence in the classification for any specific set of measurements. For example, if a bone falls into the node 7 in Figure 1, the results indicate about a 75% probability that the bone is human, despite an overall model accuracy rate of 91%. Decision trees are intuitive, transparent, and easy to apply [40,41,46]. While the concept of decision trees is not new to forensic anthropology [39,40,47–50], the method remains underutilized in practice.

Another advantage to decision tree models is that they allow users to assign higher costs to certain sets of misclassifications [36], in this case to the misclassification of human remains as nonhuman. In forensic anthropology, misclassifying human remains as nonhuman could prevent decedent identification, leaving family members without closure and impeding possible criminal investigations. In contrast, the biggest cost of misclassifying a nonhuman element as human is the unnecessary expenditure of time and resources spent in securing a scene and contacting an expert for final determination. The decision trees presented here assist in reducing the possibility of both of these scenarios. A death investigator called to a scene with a bone could have the decision tree printed on a single sheet of paper (or access it via the OsteoID website on their smartphone) and, using a tape measure, can easily follow the branches of the tree for a preliminary assessment of human versus nonhuman. Because of the integrated misclassification costs, the trees are more likely to incorrectly assign a nonhuman bone as human than vice versa; thus, the result is conservative and anything close to matching human form will be treated as if it is human and of forensic significance until determined otherwise (ideally by a trained forensic anthropologist). At the same time, resources are not wasted on scenes containing remains that are clearly not human. Thus, the models presented here can act as a triaging tool.

While some may argue that all bones discovered should be assessed by a forensic anthropologist, this is not realistic and does not represent current practice. Forensic anthropologists typically receive elements that are believed to possibly be human. Those remains that the finder, law enforcement agent, or those consulted by the law enforcement agent (including physicians and veterinarians) deem as not human are frequently not referred to medicolegal agencies or forensic anthropologists. If referred to medicolegal agencies, their non-anthropological personnel may also determine that the remains received are not human and not worth consulting with a forensic anthropologist. Resources, such as the models and web tool presented here, can assist these individuals who are already undertaking these triaging roles to make more informed decisions. If the decision trees, discriminant functions, visual comparison with the web tool images and/or context of the remains suggest that they may be of human origin, the medicolegal agency and forensic anthropologist should be consulted for final determinations. The forensic anthropologist, in turn, may find these resources useful in supporting their designations or confirming the particular faunal species (discussed below).

Not surprisingly, the most accurate human versus nonhuman functions and decision trees include measurements from multiple regions of the bone, which may not be possible in cases involving fragmented remains. Consequently, the use of only specific bone regions was tested as part of this study for application to larger bone fragments. Univariate analyses were performed on maximum lengths to reflect cases in which erosion to the epiphyses could affect proximal and distal elements. Models were created from only the distal measurements (width and depth) and from only the midshaft measurements (maximum and minimum diameters) for use in cases limited to these fragmented regions. The length and distal epiphyseal region-specific analyses produced higher accuracy rates than the midshaft measurements (except for the ulna). This is expected given that maximum length and distal width were commonly the most important variables in the more inclusive models. For the femoral decision tree, despite inputting all six variables, the tree output only used maximum length and was able to correctly classify over 96% of the total sample and over 99% of the human sample. The region-specific discriminate functions developed per bone (Supplemental Table S12) produced accuracy rates above 85% for all functions except the humeral midshaft (67.1%). These results are slightly higher than the region-

specific DFA results presented by Saulsman et al. [14]. While the results suggest that these models may be useful tools when assessing fragmented remains as human or nonhuman, caution is still warranted given that classification rates are only moderately high, and additional evidence (e.g., presence of morphological features, application of a second method) should be provided to support the conclusion. Saulsman and colleagues [14] also warn against estimating the midshaft location on humeral fragments because deviations 2 cm above or below the actual midshaft location significantly altered their classification rates; results from femoral and tibial deviations were more robust. Application of the models to burned fragments must also consider the possibility of bone shrinkage with the thermal modification [51].

The most conservative approach for assessing the human origin of skeletal remains using osteometrics would be to compare specimen measurements with the minimum, maximum, and 95% confidence intervals for human remains presented in Table 2 and at least preliminarily consider anything that falls within that range, or very close to that range, as potentially human pending further analysis. OsteoID [43] will return images of human bones if the input measurements fall anywhere within the min/max or standard deviation ranges compiled from the sample of >2700 individuals. Practitioners must always consider the small possibility that their unknown specimen can be an outlier, perhaps lying at the extremes of the human distribution which may not have been captured in this study. Pathological conditions that affect body size (e.g., dwarfism, gigantism, etc.), although rare, could also affect results [52,53].

In highly fragmented or taphonomically-modified remains, morphometric and visual assessments may not be applicable. Other evidence, such as cortical bone thickness and trabecular bone density may be factored into the decision [4,54,55], although research by Rerolle et al. [56] suggests that corticomedullary index may not be as distinctive in humans as previously suggested. Several papers state that nutrient foramen location and morphology can assist in human versus nonhuman distinctions [57,58]. Microscopic (histomorphological) or molecular methods can also be utilized [59–63] to determine human origin, but they require greater expertise and specialized equipment, are more time intensive, and are destructive to the specimen [3]. Even histomorphological techniques cannot provide 100% accuracy in distinguishing human from nonhuman species, with certain faunal species (e.g., large mammals) and bone types (e.g., presence of only Haversian bone) shown to be particularly problematic [60]. Publications also differ on opinions of the use of osteon circularity in determining human origin of bone [62,63].

4.2. Species Identification

The quantitative methods of species identification were less successful than those assessing human origin. While these results are likely impacted by uneven sample sizes across the 28 species, they also reflect morphological and size similarity between some species. For example, brown bear and black bear long bones are morphologically similar [41,64–67], especially as represented by these few basic measurements; thus, small brown bears and large black bears may be misidentified. Sheep and goat long bones are also difficult to differentiate [29,68]. Domestic dogs pose many issues, not just because of their similarity to other canids included in this study (e.g., coyotes and wolves) [69,70] but also because of their high degree of variability in both morphology and size [71,72]. The DFA species classification rates were significantly higher than chance, but the probability of species misidentification remains relatively high. The application of a discriminant function to classify an unknown specimen into one of 28 groups would also be impractical to do by hand, thereby requiring computer usage. Ultimately, practitioners must rely on visual comparisons of more subtle morphological differences in making the final faunal species designations.

In facing these challenges of species identification, the OsteoID website [43] is particularly useful. Users can input basic measurements to narrow down the potential species and are presented with photographic images of the possible identifications.

Thus, the measurements are used as a filtering tool, but the final identification is still based on visual comparison. With the use of visual comparisons, OsteoID can be used for identifying fragmented elements. Supplemental resources provided on the website can also be utilized in skeletal identifications, such as access to the metric database, a link to this publication and associated Supplementary Materials, 3D scans of numerous elements, and lists of other useful texts and websites. Photographs of additional elements (e.g., carpals) not included in the web tool are provided and will be continually updated. The web tool can easily be modified if future minimum/maximum values need revision. There is also the possibility of expanding the database and web tool to include additional species/specimens in the future.

As an online, searchable, comparative osteology collection that includes photographs, data, and 3D scans, OsteoID [43] provides forensic anthropologists with a centralized location for free resources to facilitate skeletal species identification. Practitioners with less zooarchaeological training or lacking access to physical comparative collections will benefit most from these resources when determining faunal species. The web tool and online resources can be accessed from smart phones and other devices while at the scene. With the download of free third-party applications, even the 3D bone models can be viewed on smart phones. The 3D models also can be downloaded and 3D printed to create comparative collections. Beyond forensic anthropologists, forensic pathologists, medical examiners, coroners, crime scene and death investigators, and law enforcement personnel may find OsteoID useful when making preliminary assessments. In situations where scene personnel have reason to believe that remains are nonhuman and typically would have dismissed the remains as not forensically significant, they can use the OsteoID resources to visually confirm that the morphology is not consistent with a human and perhaps find a faunal species match. In cases in which there is any possibility that remains are human, expert opinions should still be obtained. Modified remains or those that are more diagnostically difficult will require a forensic anthropologist's expertise, but OsteoID can reduce time and cost expenditures for diagnostically nonhuman remains. Bioarchaeologists, zooarchaeologists, veterinarians, and biologists may also find the OsteoID web tool and resources useful, and the general public may find interest in learning more about remains encountered. Presently, there are multiple social media groups where individuals post their skeletal finds and group participants provide species identifications. Given that OsteoID is publicly available, it contains multiple disclaimers urging anyone with remains that could potentially be human to leave them in situ and to contact local authorities. Finally, the photographs and 3D scans made available via the website can be used to train students in comparative osteology and the data may be used by researchers in other studies.

4.3. Limitations and Future Directions

Given that all forensic anthropologists rely partly on bone form (i.e., size and general shape) when assessing human origin, using bone metrics to create a quantitative classification method seems simple and logical. However, our study illustrates several challenges to this work. Firstly, it is difficult to find measurements that can be collected consistently across diverse species and bones. Limiting our measurements to maximum lengths, breadths, and depths allowed us to increase the range of animals and skeletal elements in our dataset for pooled analyses, but it excludes aspects of discrete morphological features used in visual assessments of species identification. While the general morphometric variables were able to successfully differentiate human from nonhuman remains (similar to the results of Saulsman et al. [14]), visual assessments that consider specific bone features are necessary for accurate faunal species identification.

Because the methods developed here are dependent on size and epiphyseal breadths, only skeletally mature specimens could be included in quantitative analyses (and resultant functions and models are only applicable to skeletally mature specimens). At least partial fusion of both the proximal and distal epiphyses should be observed prior to utilizing the discriminant functions or decision trees. Skeletally mature specimens

of certain species can be hard to locate, especially domesticated species which may be butchered as juveniles [73]. The species curated at museums vary and again tend not to focus on domesticated species or may not curate full skeletons, especially for larger mammals where space becomes a challenge.

Unequal sample sizes from different species could have biased our classification results, particularly with human versus nonhuman analyses. Although a high degree of faunal variation is captured in the pooled nonhuman sample, there is a smaller representation of some of the largest mammalian species. Given that humans also have relatively large body sizes, this may be driving some of the classification bias, as the models may be more likely to classify all large bones (human or nonhuman) as human given the large human sample sizes. Indeed, larger animals such as moose, brown bear, horse, cow and elk were more commonly misclassified as human, which could explain the relatively higher human and lower nonhuman classification rates in the discriminant functions. Misclassifying some of these species elements as human instead of nonhuman in preliminary forensic contexts is less costly than erroneously classifying human elements as nonhuman; following the preliminary human classification, a forensic anthropologist would then be consulted for a more formal assessment that would identify the error.

The smaller sample sizes in some nonhuman species are also less likely to capture the true population size variation and thus impact DFA species classifications. The human sample size, however, which is of greatest forensic significance, is sufficiently large, and the nonhuman sample sizes exceed those of previous publications [14]. Furthermore, not all measurements were available for all specimens. Data obtained from the literature frequently had some but not all the study measurements, meaning that in the DFAs, many of those cases were excluded.

The species included in the metric database are not exhaustive, and it is unclear how a specimen from an excluded species would classify. This study was limited to species commonly encountered in North America that were accessible at collections but does not include, for example, marine mammals. Further validation of the developed methods is needed, and if more data can be collected from additional species and specimens, revised models may be more appropriate. Future data collection for human versus nonhuman determinations should focus on adding greater samples of larger-bodied mammals. While increased samples of larger-bodied fauna may decrease model accuracy rates, it is possible that the models may still be able to confidently differentiate human from nonhuman specimens given the distinct functional anatomy of humans [3,74,75].

Preliminary analyses using a subsample of the humeral and femoral data suggest that machine learning and random forest models may be able to further increase morphometric classification rates for human versus nonhuman designations and species assignments [76]. Random forest models are a machine learning approach in which numerous decision trees are created from random subsamples, and their predictions are combined through averaging to produce a final classification [46–48]. This machine learning technique increases classification stability and alleviates potential issues of overfitting [58]. The downside of random forest models is their complexity. Because random forest model results are based on the combined results of hundreds or thousands of trees, there is no final model/tree that can be presented or applied to cases [46]. This ensemble approach is considered a "black box" method [41] meaning that it is mathematically complex and difficult to understand and explain in terms of application [77], which can be a disadvantage in court testimony. Furthermore, for broad application, a software program would need to be created to run the random forest models with new unknown specimens.

5. Conclusions

The tools presented in this study do not diminish the need for forensic anthropologists. Caution must still be used given the high cost of misclassifying a human bone as nonhuman, and forensic anthropologists or other experts should be consulted in situations where there is any possibility that remains may be human. Still, the resources developed and

provided here may be used to preliminarily assess whether remains are potentially human and determine the number of resources to expend on a found bone (e.g., whether or not a scene needs to be preserved, etc.). Forensic anthropologists or other medicolegal personnel can use the resources to support classifications and faunal species identifications. These resources may also be beneficial to other disciplines where skeletal remains are encountered or training in comparative osteology is beneficial, including wildlife forensics, bioarchaeology, zooarchaeology, veterinary medicine, and biology.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biology11010025/s1. Figure S1: Human versus nonhuman decision tree derived from all available measurements and a pooled-bone sample. Figure S2: Human versus nonhuman decision tree derived from only distal bone measurements and a pooled-bone sample. Figure S3: Human versus nonhuman decision tree derived from only midshaft measurements and a pooled-bone sample. Figure S4: Human versus nonhuman decision tree derived from only maximum length measurements using a pooled-bone sample. Figure S5: Human versus nonhuman decision tree for the humerus, derived from all available measurements. Figure S6: Human versus nonhuman decision tree for the femur, derived from all available measurements. Figure S7: Human versus nonhuman decision tree for the radius, derived from all available measurements. Figure S8: Human versus nonhuman decision tree for the tibia, derived from all available measurements. Figure S9: Human versus nonhuman decision tree for the ulna, derived from all available measurements. Table S1: Descriptive statistics for humeral measurements collected by species. Table S2: Descriptive statistics for femoral measurements collected by species. Table S3: Descriptive statistics for radial measurements collected by species. Table S4: Descriptive statistics for radio-ulnar measurements collected by species. Table S5: Descriptive statistics for ulnar measurements collected by species. Table S6: Descriptive statistics for tibial measurements collected by species. Table S7: Descriptive statistics for fibular measurements collected by species. Table S8: Descriptive statistics for scapular measurements collected by species. Table S9: Descriptive statistics for sacral measurements collected by species. Table S10: Descriptive statistics for pelvic measurements collected by species. Table S11: Descriptive statistics for fused metapodial measurements collected by species. Table S12: Select discriminant functions for human versus nonhuman classification.

Author Contributions: Conceptualization, H.M.G., R.D. and S.B.S.; methodology, H.M.G., R.D. and S.B.S.; data collection and processing, H.M.G., R.D., S.B.S., M.S.L., M.M., N.S. and N.K.; formal analysis, H.M.G.; software, H.M.G.; data curation, H.M.G. and R.D.; writing—original draft preparation, H.M.G., R.D., S.B.S.; writing—review and editing, H.M.G., R.D., S.B.S., M.S.L., M.M., N.S. and N.K.; supervision, H.M.G. and S.B.S.; funding acquisition, H.M.G., R.D. and S.B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institute of Justice, grant number NIJ 2018-DU-BX-0229. Opinions expressed herein do not necessarily represent the official position or policies of the U.S. Department of Justice or the National Institute of Justice.

Institutional Review Board Statement: All data used in this study were collected in compliance with the Declaration of Helsinki and the federal regulations for the protection of human subjects in research at 45 CFR 46 from postmortem skeletal collections curated at established museums or institutions.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available via the OsteoID website (www.boneidentification.com (accessed on 24 December 2021) → Additional Resources), as well as on Dryad (doi:10.5061/dryad.73n5tb2z0).

Acknowledgments: We would like to thank all the museum curators that facilitated access to collections. A special thanks to the following individuals who contributed their time or data to the project: Andrea Clendaniel, Elizabeth Dougher, Alexandra Klales, Michael Kenyhercz, Dennis Dirkmaat, Julie Meachen, Chelsea Cataldo-Ramirez, and Christopher Milensky. Finally, thank you to Erin Menardi, the Web Developer who collaborated with the PIs in creating this online tool.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pokines, J.T. Identification of nonhuman remains received in a medical examiner setting. *J. Forensic Identif.* **2015**, *65*, 223–246.
2. Bass, W.M. *Human Osteology—A Laboratory and Field Manual*, 4th ed.; Missouri Archaeological Society: Columbia, MO, USA, 1995; p. 309.
3. Christensen, A.M.; Passalacqua, N.V.; Bartelink, E.J. *Forensic Anthropology: Current Methods and Practices*, 2nd ed.; Academic Press: London, UK, 2019; pp. 104–123.
4. Franklin, D.; Marks, M.M. Species: Human versus nonhuman. In *Encyclopedia of Forensic Sciences*, 2nd ed.; Siegel, J.A., Saukko, P.J., Eds.; Elsevier: London, UK, 2013; Volume 1, pp. 28–33.
5. Adams, B.; Crabtree, P. *Comparative Skeletal Anatomy—A Photographic Atlas for Medical Examiners, Coroners, Forensic Anthropologists, and Archeologists*; Human Press: Totowa, NJ, USA, 2008.
6. Adams, B.; Crabtree, P. *Comparative Osteology—A Laboratory and Field Guide of Common North American Animals*; Elsevier: Oxford, UK, 2012.
7. France, D. *Human and Nonhuman Bone Identification: A Color Atlas*; CRC: Boca Raton, FL, USA, 2008.
8. France, D. *Human and Nonhuman Bone Identification: A Concise Field Guide*; CRC: Boca Raton, FL, USA, 2011.
9. France, D. *Comparative Bone Identification—Human Subadult to Nonhuman*; CRC: Boca Raton, FL, USA, 2017.
10. Hillson, S. *Mammal Bones and Teeth—An Introductory Guide to Methods of Identification*; Routledge: New York, NY, USA, 2009.
11. Beisaw, A.M. *Identifying and Interpreting Animal Bones—A Manual*; Texas A&M University Press: College Station, TX, USA, 2013.
12. Broughton, J.M.; Miller, S.D. *Zooarchaeology and Field Ecology—A Photographic Atlas*; University of Utah Press: Salt Lake City, UT, USA, 2016.
13. Olsen, S.J. *Mammal Remains from Archaeological Sites*; Papers of the Peabody Museum of Archaeology and Ethnology, President and Fellows of Harvard College: York, PA, USA, 1964; Volume 56.
14. Saulsman, B.; Oxnard, C.E.; Franklin, D. Long bone morphometrics for human from non-human discrimination. *Forensic Sci. Int.* **2010**, *202*, 110.e1–110.e5. [CrossRef] [PubMed]
15. Von den Dreisch, A. *A Guide to the Measurement of Animal Bones from Archaeological Sites: As Developed by the Institut fur Palaeoanatomie, Domestikationsforschung und Geschichte der Tiermedizin of the University of Munich*; Peabody Museum Press: Cambridge, MA, USA, 1976; Volume 1.
16. Eerkens, J.W.; Darwent, C.M.; Campbell-Grey, A.; Wildt, J.; Wheelis, B. Meaner than a junkyard dog: Osteological and isotopic insights into three canid burials from 1890s San Jose, California. *SCA Proc.* **2019**, *33*, 211–222.
17. Davis, S.J.M. Measurements of a group of adult female Shetland sheep skeletons from a single flock: A baseline for zooarchaeologists. *J. Archaeol. Sci.* **1996**, *23*, 593–612. [CrossRef]
18. Meachen, J.A.; Samuels, J.X. Evolution in coyotes (*Canis latrans*) in response to the megafaunal extinctions. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4191–4196. [CrossRef]
19. Pitakarnnop, T.; Buddhachat, K.; Euppayo, T.; Kriangwanich, W.; Nganvongpanit, K. Feline (*Felis catus*) skull and pelvic morphology and morphometry: Gender-related difference? *Anat. Histol. Embryol.* **2017**, *46*, 294–303. [CrossRef] [PubMed]
20. Knoest, J.J.T. On the Morphology of the Domesticated Dog in Medieval Norway: An Osteometric Study. Master's Thesis, University of Bergen, Bergen, Norway, December 2015.
21. Wagner, K. *Rezente Hunderassen. Eine Osteologische Untersuchung*; Antropologische Abteilung: Oslo, Norway, 1930.
22. Boonsri, B.; Pitakarnnop, T.; Buddhacat, K.; Changtor, P.; Nganvongpanit, K. Can feline (*Felis catus*) flat and long bone morphometry predict sex or skull shape? *Anat. Sci. Int.* **2019**, *94*, 245–256. [CrossRef]
23. Zeder, M.A. A metrical analysis of a collection of modern goats (*Capra hircus aegagrus* and *C. h. hircus*) from Iran and Iraq: Implications for the study of Caprine domestication. *J. Archaeol. Sci.* **2001**, *28*, 61–79. [CrossRef]
24. Silvia, W.J.; Peterson, R.O.; Vucetich, J.A.; Silvia, W.F.; Silvia, A.W. Variation in metatarsal morphology among subgroups of North American Moose (*Alces alces*). *Alces J. Devoted Biol. Manag. Moose* **2014**, *50*, 159–170.
25. Saulsman, B. Long Bone Morphometrics for Human from Non-human Discrimination. Master's Thesis, The University of Western Australia, Perth, Australia, 2010.
26. Guintard, C.; Lallemand, M. Osteometric study of metapodial bones in sheep (*Ovis aries*, L. 1758). *Ann. Anat.* **2003**, *185*, 573–583. [CrossRef]
27. Doyle, S.K. Scaling Patterns and Ecological Correlates of Postcranial Skeletal Robusticity in *Canis* and *Ursus*: Implications for Human Evolution. Ph.D. Thesis, Duke University, Durham, NC, USA, 2009.
28. Gilbert, B.M. *Mammalian Osteology*; Missouri Archaeological Society: Columbia, MO, USA, 1990.
29. Salvagno, L.; Albarella, U. A morphometric system to distinguish sheep and goat postcranial bones. *PLoS ONE* **2017**, *12*, e0178543. [CrossRef] [PubMed]
30. Thomas, R.; Holmes, M.; Morris, J. "So bigge as bigge may be": Tracking size and shape change in domestic livestock in London (AD 1220–1900). *J. Arcchaeol. Sci.* **2013**, *40*, 3309–3325. [CrossRef]
31. Pfeiffer, S. Carpine skeletons can be "herd". *SCA Proc.* **2013**, *27*, 237–241.
32. Hunt, D.; Dollas, N.; DiMichele, D.; Robert, J. Terry Anatomical Human Skeletal Collection database. National Museum of Natural History. *Dataset* **2019**. [CrossRef]
33. Auerbach, B.M. Goldman Osteometric Data Set. Available online: https://web.utk.edu/~{}auerbach/GOLD.htm (accessed on 1 October 2019).

34. Samuels, J.X.; Meachen, J.A.; Sakai, S.A. Postcranial morphology and the locomotor habits of living and extinct carnivorans. *J. Morphol.* **2013**, *274*, 121–146. [CrossRef] [PubMed]
35. Machuca, C.; Vettore, M.V.; Krasuka, M.; Baker, S.R.; Robinson, P.G. Using classification and regression tree modelling to investigate response shift patterns in dentine hypersensitivity. *BRC Med. Res. Methodol.* **2017**, *17*, 120. [CrossRef]
36. *IBM SPSS Decision Trees 21*; IBM Corporation: Armonk, NY, USA, 2012.
37. Bartholdy, B.P.; Sandoval, E.; Hoogland, M.L.P.; Schrader, S.A. Getting rid of dichotomous sex estimations: Why logistic regression should be preferred over discriminant function analysis. *J. Forensic Sci.* **2020**, *65*, 1685–1690. [CrossRef] [PubMed]
38. Walker, P.L. Sexing skulls using discriminant function analysis of visually assessed traits. *Am. J. Phys. Anthr.* **2008**, *136*, 39–50. [CrossRef]
39. Stevenson, J.C.; Mahoney, E.R.; Walker, P.L.; Everson, P.M. Prediction of sex based on five skull traits using decision analysis (CHAID). *Am. J. Phys. Anthropol.* **2009**, *139*, 434–441. [CrossRef] [PubMed]
40. Langley, N.R.; Dudzik, B.; Cloutier, A. A decision tree for nonmetric sex assessment from the skull. *J. Forensic Sci.* **2018**, *63*, 31–37. [CrossRef] [PubMed]
41. Nikita, E.; Nikitas, P. Data mining and decision trees. In *Statistics and Probability in Forensic Anthropology*; Obertova, Z., Stewart, A., Cattaneo, C., Eds.; Academic Press: London, UK, 2020; pp. 88–105.
42. Feldesman, M.R. Classification trees as an alternative to linear discriminant analysis. *Am. J. Phys. Anthr.* **2002**, *119*, 257–275. [CrossRef] [PubMed]
43. OsteoID. Available online: www.boneidentification.com (accessed on 29 October 2021).
44. Morphosource OsteoID Project. Available online: https://www.morphosource.org/projects/000364427?locale=en (accessed on 27 November 2021).
45. Garvin, H.M.; Dunn, R.; Sholts, S. Postcranial osteometric data from human and 27 North American faunal species, Dryad. *Dataset* **2021**. [CrossRef]
46. Algee-Hewitt, B.; Coelho, C.; Navega, D.; Cunha, E. New and established methods for the quantification of cranial variation for forensic casework. In *Statistics and Probability in Forensic Anthropology*; Obertova, Z., Stewart, A., Cattaneo, C., Eds.; Academic Press: London, UK, 2020; pp. 230–235.
47. Nikita, E.; Nikitas, P. On the use of machine learning algorithms in forensic anthropology. *Leg. Med.* **2020**, *47*, 101771. [CrossRef] [PubMed]
48. Hefner, J.T.; Ousley, S.D. Statistical classification methods for estimating ancestry using morphoscopic traits. *J. Forensic Sci.* **2014**, *59*, 883–890. [CrossRef] [PubMed]
49. Navega, D.; Coelho, C.; Vicente, R.; Ferriera, M.T.; Wasterlain, S.; Cunha, E. AncesTrees: Ancestry estimation with randomized decision trees. *Int. J. Leg. Med.* **2015**, *129*, 1145–1153. [CrossRef] [PubMed]
50. Savall, F.; Faruch-Bilfeld, M.; Dedouit, F.; Sans, N.; Rousseau, H.; Rouge, D.; Telmon, N. Metric sex determination using the human coxal bone on a virtual sample using decision trees. *J. Forensic Sci.* **2015**, *60*, 1395–1400. [CrossRef]
51. Ubelaker, D.H. The forensic evaluation of burned skeletal remains: A synthesis. *Forensic Sci. Int.* **2009**, *183*, 1–5. [CrossRef] [PubMed]
52. Frayer, D.W.; Macchiarelli, R.; Mussi, M. A case of chondrodystrophic dwarfism in the Italian late Upper Paleolithic. *Am. J. Phys. Anthr.* **1988**, *75*, 549–565. [CrossRef]
53. Mulhern, D.M. A probable case of gigantism in a fifth dynasty skeleton from the Western Cemetery at Giza, Egypt. *Int. J. Osteoarchaeol.* **2005**, *15*, 261–275. [CrossRef]
54. Croker, S.L.; Clement, J.G.; Donlon, D. A comparison of cortical bone thickness in the femoral midshaft of humans and two non-human mammals. *HOMO—J. Comp. Hum. Biol.* **2009**, *60*, 551–565. [CrossRef]
55. Croker, S.L.; Reed, W.; Donlon, D. Comparative cortical bone thickness between the long bones of humans and five common non-human mammal taxa. *Forensic Sci. Int.* **2016**, *260*, 104.e1–104.e17. [CrossRef] [PubMed]
56. Rerolle, C.; Saint-Martin, P.; Dedouit, F.; Rousseau, H.; Telmon, N. Is the corticomedullary index valid to distinguish human from nonhuman bones: A multislice computed tomography study. *Forensic Sci. Int.* **2013**, *231*, 406.e1–406.e5. [CrossRef]
57. Corrieri, B.; Marquez-Grant, N. Using nutrient foramina to differentiate human from non-human long bone fragments in bioarchaeology and forensic anthropology. *HOMO—J. Comp. Hum. Biol.* **2019**, *70*, 255–268. [CrossRef] [PubMed]
58. Johnson, V.; Beckett, S.; Marquez-Grant, N. Differentiating human versus non-human bone by exploring the nutrient foramen: Implications for forensic anthropology. *Int. J. Leg. Med.* **2017**, *131*, 1757–1763. [CrossRef] [PubMed]
59. Brits, D.; Steyn, M. A histomorphological analysis of human and non-human femora. *Int. J. Leg. Med.* **2014**, *128*, 369–377. [CrossRef]
60. Hillier, M.L.; Bell, L.S. Differentiating human bone from animal bone: A revie of histological methods. *J. Forensic Sci.* **2007**, *52*, 249–263. [CrossRef] [PubMed]
61. Mulhern, D.M.; Ubelaker, D.H. Differences in osteon banding between human and nonhuman bone. *J. Forensic Sci.* **2001**, *46*, 220–222. [CrossRef]
62. Crescimanno, A.; Stout, S.D. Differentiating fragmented human and nonhuman long bone using osteon circularity. *J. Forensic Sci.* **2012**, *57*, 287–294. [CrossRef]
63. Dominguez, V.M.; Crowder, C.M. The utility of osteon shape and circularity for differentiating human and non-human Haversian bone. *Am. J. Phys. Anthr.* **2012**, *149*, 84–91. [CrossRef] [PubMed]

64. Bogner, E. Differentiating Black Bears (*Ursus americanus*) and Brown Bears (*U. arctos*) using Linear Tooth Measurements and Identification of Ursids from Oregon Caves National Monument. Master's Thesis, East Tennessee State University, Johnson City, TN, USA, May 2019.
65. Kantelis, T.M. Black Bears (*Ursus americanus*) versus Brown Bears (*U. arctos*): Combining Morphometrics and Niche Modeling to Differentiate Species and Predict Distributions Through Time. Master's Thesis, East Tennessee State University, Johnson City, TN, USA, May 2017.
66. Lariviere, S. Ursus americanus. *Mamm. Species* **2001**, *64*, 1–11. [CrossRef]
67. Pasitschniak-Arts, M. Ursus arctos. *Mamm. Species* **1993**, *439*, 1–10. [CrossRef]
68. Boessneck, J. Osteological differences between sheep (*Ovis aries* Linne). In *Science in Archaeology: A Comprehensive Survey of Progress and Research*, 2nd ed.; Browthwell, D., Higgs, E., Eds.; Thames & Hudson: London, UK, 1969; pp. 311–358.
69. Welker, M.H.; Byers, D.A.; McClure, S.B. "I wanna be your dog": Evaluating the efficacy of univariate and multivariate methods for differentiating domestic and wild canids in North America. *Int. J. Osteoarchaeol.* **2021**, *31*, 196–206. [CrossRef]
70. Wayne, R.K.; Geffen, E.; Girman, D.J.; Koepfli, K.P.; Lau, L.M.; Marshall, C.R. Molecular systematics of the Canidae. *Syst. Biol.* **1997**, *46*, 622–653. [CrossRef]
71. Bannasch, D.L.; Baes, C.F.; Leeb, T. Genetic variants affecting skeletal morphology in domestic dogs. *Trends Genet.* **2020**, *36*, 598–609. [CrossRef] [PubMed]
72. Frynta, D.; Baudysova, J.; Hradcova, P.; Faltusova, K.; Kratochvil, L. Allometry of sexual size dimorphism in domestic dog. *PLoS ONE* **2012**, *7*, e46125. [CrossRef] [PubMed]
73. Holmes, M.; Hamerow, H.; Thomas, R. Close companions? a zooarchaeological study of the human-cattle relationship in Medieval England. *Animals* **2021**, *11*, 1174. [CrossRef] [PubMed]
74. Preuschoft, H. Mechanisms for the acquisition of habitual bipedality: Are there biomechanical reasons for the acquisition of upright bipedal posture? *J. Anat.* **2004**, *204*, 363–384. [CrossRef] [PubMed]
75. Kardong, K.V. *Vertebrates: Comparative Anatomy, Function, Evolution*, 6th ed.; McGraw-Hill: New York, NY, USA, 2009; pp. 348–371.
76. Garvin, H.M.; Kenyhercz, M.W.; Dunn, R.; Sholts, S.; Litten, M.S. A Metric Approach to Human and Non-Human Species Identification of Skeletal Remains using Machine Learning. In Proceedings of the American Academy of Forensic Sciences 73rd Annual Scientific Meeting, Virtual, 15–19 February 2021; p. 80.
77. Loyola-Gonzalez, O. Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access* **2019**, *7*, 154096–154113. [CrossRef]

biology

MDPI

Article

Providing a Forensic Expert Opinion on the "Degree of Force": Evidentiary Considerations

Hans H. de Boer [1,*,†], Charles E. H. Berger [2,3,†] and Soren Blau [1,†]

[1] Victorian Institute of Forensic Medicine/Department of Forensic Medicine, Monash University, Southbank, VIC 3006, Australia; soren.blau@vifm.org
[2] Institute of Criminal Law and Criminology, Leiden University, P.O. Box 9500, 2300 RA Leiden, The Netherlands; C.Berger@nfi.nl
[3] Netherlands Forensic Institute, Ministry of Justice and Security, Government of The Netherlands, 2511 DP The Hague, The Netherlands
* Correspondence: hans.de.boer@vifm.org
† All authors contributed equally.

Simple Summary: When giving evidence in court, forensic pathologists and anthropologists are often asked for their opinion on the amount, or degree of force required to cause a specific injury. Such 'degree of force' questions are considered difficult, if not impossible to answer due to many theoretical and practical issues. This paper explores these issues and provides a possible solution. First, the logical underpinnings of the question on the 'degree of force' are explored. Then the experimental research on 'degree of force' is reviewed and the limitations with applying this research to everyday forensic casework are discussed. In the second part of the paper, it is argued that these limitations do not, however, mean that a forensic pathologist or anthropologist cannot add anything of value to the discussion. The application of Bayes' theorem helps to circumvent many of the problems. The final part of the paper is dedicated to a detailed discussion of how it can be applied to the issue of 'degree of force'.

Abstract: Forensic pathologists and anthropologists are often asked in court for an opinion about the degree of force required to cause a specific injury. This paper examines and discusses the concept of 'degree of force' and why it is considered a pertinent issue in legal proceedings. This discussion identifies the implicit assumptions that often underpin questions about the 'degree of force'. The current knowledge base for opinions on the degree of force is then provided by means of a literature review. A critical appraisal of this literature shows that much of the results from experimental research is of limited value in routine casework. An alternative approach to addressing the issue is provided through a discussion of the application of Bayes' theorem, also called the likelihood ratio framework. It is argued that the use of this framework makes it possible for an expert to provide relevant and specific evidence, whilst maintaining the boundaries of their field of expertise.

Keywords: degree of force; skeletal trauma; forensic pathology; forensic anthropology; review; evidence; opinion; likelihood ratio; Bayes' theorem

Citation: de Boer, H.H.; Berger, C.E.H.; Blau, S. Providing a Forensic Expert Opinion on the "Degree of Force": Evidentiary Considerations. *Biology* **2021**, *10*, 1336. https://doi.org/10.3390/biology10121336

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 1 November 2021
Accepted: 14 December 2021
Published: 16 December 2021

1. Introduction

> *" . . . force alone is woefully inadequate and often (particularly in a legal environment) misleading in describing an impact".* [1], p. 283

The concept that skeletal trauma occurs when a force exceeds the strength or maximum threshold of bone elasticity is well established [2,3]. In forensic pathology and anthropology, descriptions of the application of a force to the body are typically divided into three groups of causation: blunt force, sharp force and high or low energy ballistic force. While the potential results of these forces on the human body have been well documented in the biomechanical [3–5] and forensic medical literature [1,6,7], correlating the amount

of force applied to the body to a specific injury or fracture outcome has proven more difficult. Nonetheless, when providing expert opinion in court about skeletal injury, forensic pathologists and anthropologists are often asked their opinion about the cause and specifically, the amount or the 'degree of force' that was required.

The relationship between injury morphology and applied force is complex, and opinions on the 'degree of force' are therefore fraught with difficulties. This paper provides an overview of the concept of 'degree of force' in forensic pathology and anthropology and in doing so, provides an aid for practitioners when giving evidence on this issue. While the paper focuses on skeletal injuries, much of the discussion also applies to the same issue when interpreting soft tissue injuries.

2. Why Is the 'Degree of Force' Considered to Be Important?

In order to understand why the issue of 'degree of force' appears to be so pertinent in criminal cases it is necessary to reflect briefly on the purpose of a criminal court proceeding, and the role of the expert witness. Although differences between jurisdictions exist, a criminal court proceeding typically aims to determine whether enough evidence is available to convict the defendant for the alleged crime. This process is traditionally dialectic, with prosecution and defence both trying to convince the trier-of-fact of their respective positions, usually by presenting evidence. The opinion of the expert witness, like other evidence, can assist the trier-of-fact in weighing the competing positions of prosecution and defence.

Within this context, opinions on the 'degree of force' have been considered useful to help the trier-of-fact to reconstruct the events that led up to and resulted in death. In other words, such opinions are intended to help the trier-of-fact choose between various scenarios. The high frequency with which forensic pathologists [8–11]; forensic anthropologists, and forensic physicists [12] are confronted with the question suggests that such an opinion is considered particularly helpful by the court. This perceived value of opinions on the 'degree of force' appears to be based on three assumptions, namely that proportional relationships exist between:

1. The intent of an offender/assailant, and the amount of force they use;
2. The amount of force that an offender/assailant uses, and the amount of force that is actually transferred to the body of the decedent;
3. The amount of force that is applied on the body of the decedent, and the severity of injury.

If all three assumptions are valid, the conjecture is that knowing the 'degree of force' may help to differentiate between intentional or accidental injuries, and therefore, help to conclude if a crime was actually committed. Furthermore, since the seriousness of a crime ordinarily influences sentencing decisions [13], intent is an important aspect of culpability in many jurisdictions. An expert opinion that can inform on intent can therefore have an impact on sentencing [14].

3. Forensic Expert Responses to the Question of 'Degree of Force'

Questions relating to the 'degree of force' may be asked in various forms. Typically, however, the expert is asked the question in a simplistic form: "what degree of force is required to cause this skeletal injury?" The expert is subsequently expected to provide an estimate of that amount of force, based on the combination of observations, knowledge, and experience.

Anecdotal information, largely obtained from discussions amongst forensic experts, indicates there is variation in their responses. The general consensus is that a specific answer (i.e., including a number expressing the amount of force) cannot be provided. As a result, experts may provide a response along the lines of "I am unable to comment" or "I can comment, but without a degree of precision". Other types of responses include "the force was sufficient to result in a skeletal injury"; or "clearly there has been enough force to fracture a bone". Since such comments only reiterate the facts that are already known, it may be argued that these opinions are as uninformative as "no comment".

To simplify the issue, some experts choose to use a qualitative three- or four-point scale to describe the amount of applied force. This approach has also been described in the literature, with verbal descriptions such as "mild", "moderate" and "severe" force used by Nolan et al. [15]; and "mild", "moderate", "considerable" and "severe" force by Gilchrist et al. [16] and Sharkey et al. [17]. A definition of what these specific categories mean, or how the expert should choose between them, however, remains largely undiscussed. In a study pertaining to stab injuries, Gilchrist and colleagues [16] stated that a mild level of force would "typically" be associated with penetration of skin and soft tissue, moderate force with injuries that penetrate cartilage or rib bone, and severe force with injuries that penetrate dense bone and cause visible damage to the knife's blade. But these definitions are not generally accepted, and the limitations of using these vague and relative terms have been previously noted [8,15].

Overall, there is no consensus on how an expert should answer a question on the 'degree of force'. This lack of consensus has served as a justification for research which has sought to quantify the degree of force in various types of injuries.

4. Evidence for the Relationship between Degree of Force and Injury Outcome

A range of experimental studies have been undertaken to investigate and correlate the relationship between degree of force and injury outcome. This research ordinarily focuses on the method of injury, rather than on the type of tissue injured. For instance, research has included the investigation of degree of force and sharp force trauma involving knives [8,16,18], as well as stabbing involving other implements such as screwdrivers [19]. Such research has used pork skin [15,18] as well as synthetic materials such as foam [20,21], silicone rubbers [22,23] and modelling clay [24] as substitutes for human skin. Research investigating the relationship between degree of force and blunt force trauma has also been undertaken. This research has mostly focused on understanding the force required to cause head injuries, including brain injuries [25] and skull fractures [26]. As experimental models, researchers have used human skulls [27] as well as those of pigs [17,26] and monkeys [28], in addition to computer simulations [29].

Despite the use of these various experimental models, different anatomical parts of the body, and different types and amounts of force, the results of these experimental studies are difficult to apply to forensic casework. This shortcoming becomes more apparent when reconsidering the previously mentioned three assumptions that underpin the alleged validity of the 'degree of force' question.

Experimental research has predominantly focused on the third assumption: the relationship between the force applied to the body, and the severity of injury. Consequently, such research only addresses one part of the issue at hand. Further, the highly controlled settings typical of experiments do not (and cannot) take into account the many intrinsic and extrinsic variables that influence the relationship between applied force and injury outcome. Intrinsic variables include the sex and age of the deceased, and the specific anatomical region impacted (e.g., head vs. chest). The anatomical region, and therefore the skeletal element, is also important to consider, as different bones differ in their density, flexibility, and design (e.g., the area of impact may be buttressed by other anatomical structures) [27]. In addition to the health status which affects bone plasticity [30], individual variation in bone morphology must also be considered (e.g., skull thickness [31–33]). Overall, while the results of experimental research may be interesting as a means of demonstrating the biomechanical properties of human (and non-human) tissue, they are not directly transferable to forensic casework.

Published research focused much less on the first and second assumptions that underpin the 'degree of force' question, that is, the relationship between 'intent' and 'force used', and between 'force used' and 'force transferred'. The difference between 'force used' and 'force transferred' is an often-overlooked issue in experimental settings but is, nonetheless, relevant in forensic casework. In many instances the relationship between these two forces is not proportional. Consider, for example, a situation in which a perpetrator exerts what

may be described as a 'relatively minimal' force (e.g., a gentle, yet intentional push) which nonetheless results in what is described as 'severe trauma' (e.g., multiple comminuted fractures due to a fall from height). In other settings the relationship between 'force used' and 'force transferred' may be proportional, but there is no way of knowing the extent to which one is influenced by the other. For example, in the case where a perpetrator uses an implement that modifies the force that is used (e.g., a baseball bat, a hammer, or a knife). Extrinsic variables such as the effect of the size, shape, elasticity, and mass of the impacting implement [17,27] are important in this regard. The directionality of the impact is also of interest [12] as well as its speed (because bone is viscoelastic, that is, responds differently depending on the speed at which a load is applied). The direction-ality and speed of impact relate directly to the relative position of perpetrator and victim and the dynamics of the event. When all these variables are considered, it becomes appar-ent that the same amount of force used by a perpetrator can, depending on the circumstances, result in different amounts of force being transferred (applied) to the body of the victim.

The assumed relationship between the intent of the perpetrator and the force that is used is also rarely considered in empirical research. Although it is intuitively true that the intention to inflict grievous bodily harm results in a large amount of force being used, it is not necessarily so that unintentional behaviour results in less force. Consider, for example, scenarios of self-defence, in which forcefully fending off an attack can cause serious harm to the attacker. Moreover, one study showed that when volunteers were asked to use 'mild', 'moderate' or 'severe' force, the resultant amounts of (stabbing) force were too similar to reliably infer the 'intent' of the volunteer [15]. The sex and age of the perpetrator have been noted as important variables to consider in this [12,15]. However, these are only two of a multitude of interacting variables that may be of relevance.

Overall, while the findings from experimental research can perhaps support claims about the potential effects of force on the human body, the data seem of limited use to provide informative opinions on the 'degree of force'. It may be argued that, in fact, the results from experimental research reinforce the idea that the 'degree of force' is an issue associated with great complexity and uncertainty, while its relevance is very limited.

5. Taking a Different Approach: Applying Bayes' Theorem

Given the complexity and uncertainty associated with providing an opinion on the 'degree of force', an alternative approach is to use probability, described as "a tool to handle uncertainty" [34]. The difficulties surrounding the issue can perhaps be addressed by applying the laws of logic and probability, using Bayes' theorem. This theorem describes the logical underpinnings of the process by which probabilities are updated based on observations [35]. Many textbooks and journal articles provide introductions to Bayes' theorem, and explain why its use is the logically correct way to interpret and present forensic evidence [34,36–38]. Bayes' theorem has been applied in a range of forensic disciplines including pathology [39,40], anthropology [41,42], entomology [43], biometrics [44], and biomechanics [45], and to address different questions such as time since death [46], manner of death [47], and identification [48,49], including disaster victim identification [50–52] and missing persons investigations [53]. To date, however, Bayes' theorem has not yet been applied to address the issues associated with opinions on the 'degree of force'.

Bayes' theorem, which in forensic science is also referred to as 'the likelihood ratio framework', is best explained by the equation in odds form:

$$\frac{P(H1)}{P(H2)} \times \frac{P(E|H1)}{P(E|H2)} = \frac{P(H1|E)}{P(H2|E)}$$

With:

P(Hx) = prior probability of proposition x

P(E | Hx) = probability of the evidence E, given proposition x

P(Hx | E) = posterior probability of proposition x, i.e., given the evidence E

This can also be written as:

$$\text{prior odds} \times \text{likelihood ratio} = \text{posterior odds}$$

It should, however, be kept in mind that this equation only shows the logical relationship between the probabilities of observations and propositions. The theorem therefore remains valid in the absence of numerical data.

5.1. Prior Odds

The prior odds are given by the ratio of the probability of proposition H1 and that of H2, without considering the expert's observations, that is, the evidence (*E*). Because the prior odds are based on all information outside the expert's evidence, assessing the prior odds would take the expert outside their area of expertise.

5.2. Likelihood Ratio

To provide an opinion while staying within their area of expertise, the expert needs to focus on the likelihood ratio (LR) only. The LR is the ratio of two probabilities: the probability of their observations (*E*) given one proposition is true, and the probability of the same observations given an alternative (mutually exclusive) proposition is true. Assessing these two probabilities relies directly on the experience and expertise of the expert. This process does not necessarily imply using statistics and calculations: the same logic applies with or without the use of numerical data.

5.3. The Posterior Odds

The posterior odds take all the evidence into account: they equal the prior odds multiplied by the LR. Since the posterior odds require the prior odds, the posterior odds are also outside the forensic pathologist or anthropologist's area of expertise.

6. A Hypothetical Case Example

The utility of the application of this framework to the issue of 'degree of force' can be illustrated by the following hypothetical case. The partially skeletonized remains of an adult male were located at the bottom of a mine shaft. The individual's skull was fragmented. The remains were examined by a forensic pathologist and a forensic anthropologist. Reconstruction of the skull fragments revealed two concentric, patterned impact fractures: one in the left fronto-temporal region, and the other in the left temporo-parietal region. There was also a linear defect on the right posterior aspect of the occipital bone. In their joint report, the forensic pathologist and anthropologist concluded that these observations indicated multiple impacts, and that the cranial trauma was a reasonable cause of death. Eventually, a person was arrested in relation to the matter and the case went to trial. In court, the experts were asked their opinion on the 'degree of force' required to produce this fragmentation and patterned injury.

While the experts can try to answer this question, as previously discussed, many limitations preclude the provision of a robust opinion. Using vague terms such as 'mild', 'moderate', 'severe' and 'extreme' to describe force does not overcome these limitations. These restrictions do not, however, mean the expert cannot add anything of value to the discussion.

7. The Need for Propositions

When applying Bayes' theorem, the expert's opinion is used as evidence to help give weight to one of two propositions, most often the positions of the prosecution and defence. For instance, in the hypothetical case outlined above, the prosecution may allege that the decedent was beaten to death with a shovel and then dumped in the mine shaft. In contrast, the defence may propose that the skull fractures were the result of a fall following a verbal altercation between the decedent and the defendant. As discussed previously, information about the 'degree of force' is just an intermediate step in addressing the larger issue: which

of the two propositions is correct. If the court is focused on one specific injury and only enquires about the 'degree of force' required, these propositions are not made explicit to the expert. Consequently, the full meaning of the pathological/anthropological findings cannot be borne out.

Only when provided with propositions, can the expert provide the most relevant evidence. For instance, in the hypothetical case, the expert could clarify that the propositions provided by prosecution and defence both imply that substantial force was applied to the skull, and therefore, an opinion on the 'degree of force' is of no use to distinguish between the two propositions. Further, by focusing on 'degree of force', other observations made by the expert remain undisclosed. In the hypothetical case example such information includes the findings that the victim had a minimum number of three impacts, both sides of the skull were impacted, and that there were two patterned impression fractures and one linear fracture. These details are all potentially useful to the court proceedings, especially when the expert is provided with some case circumstances.

8. How Does the Expert Assess Evidential Strength (An LR)?

Instead of requesting an opinion on the 'degree of force', a more appropriate question for the expert may be: "to what extent do your observations support scenario A (that the decedent was assaulted with a shovel and dumped in the mineshaft) vs. scenario B (that the decedent fell into the mineshaft)?" When confronted with these propositions, the expert can apply Bayes' theorem, and therefore provide the evidential strength of their observations (an LR).

But how are experts supposed to assess an LR? Where do they get 'the numbers' from? It is important to remember that the use of probability does not imply statistics and calculations [34], and that a lack of data does not preclude the application of logic. LRs can be used qualitatively. However, the LR framework cannot mitigate gaps in scientific knowledge. If the expert thinks there is insufficient scientific knowledge to provide an opinion, it is their professional obligation to say so. In that situation the expert's opinion represents an LR of 1, which simply means that in the expert's opinion, their observations do not assist in distinguishing between the two propositions.

In the hypothetical case the observations of the two concentric, left-sided patterned impact fractures in the fronto-temporal and temporo-parietal regions, and the right-sided linear defect in the occipital region are the relevant evidence (E). The first question is, therefore, to what extent does the expert expect (or is surprised by) these observations if scenario A (H1) is true? How probable is the presence of a linear fracture when hit with a shovel? And would such an impact result in multiple concentric, patterned impact fractures? Moving to scenario B (H2), what is the probability of the observations if the deceased just fell in the shaft without being beaten? Answers to these questions rely on the expert's observations. Ideally, however, they would also be informed by some (preferably undisputed) information on the case circumstances. In this case this information would include details about the structure (walls and bottom), height, and width of the mine shaft. To obtain an LR the expert finally needs to relate the expectation for the observations under both propositions to one another, because it is their ratio that determines the evidential strength. It is important to remember that having a low expectation for the observations under one proposition does not imply support for the other proposition, since the observations could be even more improbable under the alternative proposition.

Suppose that in the hypothetical case the mine shaft was dug into soil, did not contain any rocks, and was six meters deep. In these circumstances there is a much higher expectation for the three fractures under scenario A (the decedent was assaulted with a shovel and dumped in the mineshaft) than under scenario B (the deceased just fell in the shaft without being beaten). Suppose the probability of the observations is assessed to be higher by a factor of hundreds for scenario A versus scenario B. If the opinion scale as defined in the 'Guideline for Evaluative Reporting' by the European Network of Forensic Science Institutes (ENFSI) [54] is used, LRs in the range between 100 and 1000

are represented as 'moderately strong support'. With reference to that scale, the expert would report that the observations offer 'moderately strong support' for scenario A over scenario B. The 'moderately strong support' is the qualitative LR in this example. This LR does not imply that scenario A is the most probable scenario, as other evidence (prior odds) could point to scenario B. It does, however, mean that this expert opinion offers moderately strong support for the case of the prosecution. How to best communicate (verbal) LRs is discussed in more detail in [55].

Note how the propositions enable the expert to use all their observations to answer the question, instead of focusing on one (often out of context) single element (i.e., the degree of force). This approach increases the amount of information that can be used for the opinion. Instead of being constrained to the limited empirical evidence for the relationship between force and injury morphology, the expert can now use other sources of information as well. For instance, the expert can refer to published literature which provides an evidence base for the types of skull fractures associated with different categories of trauma (e.g., [27,56]), or fracture patterns, i.e., the number, location and morphology of skull fractures in falls [57] vs. assaults [58].

Making the question explicit in the form of propositions allows the expert to provide an LR. It furthermore clarifies the issues most relevant to the court and therefore allows the expert to maximize the relevance of their evidence. Moreover, as previously discussed in various other publications dedicated to the application of Bayes' theorem in forensic science, it helps to maintain the separate roles of the trier-of-fact and the expert, and helps to interpret evidence in a logically correct way. Thus, when asked the right question, the expert can appropriately draw on their expertise and therefore, inform the court in the most meaningful way.

9. Conclusions

Questions relating to the 'degree of force' often implicitly assume that such an opinion assists the court in establishing whether an injury was caused accidentally or intentionally. As demonstrated in this paper, this assumption is flawed, since theoretical and practical limitations preclude a connection between the 'degree of force' and intent. Similar to forensic biomechanical injury assessment, providing an opinion about the 'degree of force' does not occur in a vacuum [45], that is, all lines of evidence must be considered. The use of Bayes' theorem helps to accomplish this, and therefore enables the expert to maximize the full potential of their evidence.

Author Contributions: All authors contributed equally to the paper. Conceptualization, H.H.d.B., C.E.H.B. and S.B.; writing—original draft preparation, H.H.d.B., C.E.H.B. and S.B.; writing—review and editing, H.H.d.B., C.E.H.B. and S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Porta, D.J. Biomechanics of impact injury. In *Forensic Medicine of the Lower Extremity: Human Identification and Trauma Analysis of the Thigh, Leg, and Foot*; Rich, J., Dean, D.E., Powers, R.H., Eds.; Humana Press: Totowa, NJ, USA, 2005; pp. 279–310.
2. Shore, S.W.; Unnikrishnan, G.U.; Hussein, A.I.; Morgan, E.F. Bone biomechanics. In *Orthopaedic Biomechanics*; Winkelstein, B.A., Ed.; CRC Press: Boca Raton, FL, USA, 2013.
3. Kieser, J.; Taylor, M.; Carr, D. Biomechanics of Bone and Bony Trauma. In *Forensic Biomechanics*; Wiley-Blackwell: Oxford, UK, 2013; pp. 35–70.

4. Yoganandan, N.; Pintar, F.A. Biomechanics of Temporo-Parietal Skull Fracture. *Clin. Biomech. (Bristol Avon)* **2004**, *19*, 225–239. [CrossRef]
5. Frankel, V.H.; Kaplan, D.J.; Egol, K.A. Biomechanics of Fractures. *J. Orthop. Trauma* **2016**, *30* (Suppl. 2), S2–S6. [CrossRef]
6. Love, J.C.; Wiersema, J.M. Skeletal trauma: An anthropological review. *Acad. Forensic Pathol.* **2016**, *6*, 463–477. [CrossRef] [PubMed]
7. Berryman, H.E.; Berryman, J.F.; Saul, T.B. Bone trauma analysis in a forensic setting: Theoretical basis and a practical approach for evaluation. In *Forensic Anthropology: Theoretical Framework and Scientific Basis*; Boyd, C.C., Boyd, D.C., Eds.; Wiley: Hoboken, NJ, USA, 2018; pp. 213–234.
8. Knight, B. The dynamics of stab wounds. *Forensic Sci.* **1975**, *6*, 249–255. [CrossRef]
9. Taylor, G. Paula Conlon was stabbed 11 times with high degree of force, court told. *ABC News*, 15 August 2014.
10. Barnes, M. Inquest into the death of Michael John Eddy. 2007. Available online: https://www.courts.qld.gov.au/__data/assets/pdf_file/0004/86710/cif-eddy-mj-20070212.pdf (accessed on 30 October 2021).
11. Saukko, P. *Knight's Forensic Pathology*, 4th ed.; Knight, B., Ed.; CRC Press: Boca Raton, FL, USA, 2015.
12. Siegenthaler, L.; Sprenger, F.D.; Kneubuehl, B.P.; Jackowski, C. Impact energy of everyday items used for assault. *Int. J. Leg. Med.* **2018**, *132*, 211–217. [CrossRef]
13. Anonymous. *Victorian Sentencing Manual*, 4th ed.; Judicial College of Victoria: Melbourne, Australia, 2021.
14. Sherman, L.; Neyroud, P.W.; Neyroud, E. The Cambridge Harm Index: Measuring total harm from crime based on sentencing guidelines. *Policing* **2016**, *10*, 171–183. [CrossRef]
15. Nolan, G.; Hainsworth, S.V.; Rutty, G.N. Forces generated in stabbing attacks: An evaluation of the utility of the mild, moderate and severe scale. *Int. J. Leg. Med.* **2018**, *132*, 229–236. [CrossRef]
16. Gilchrist, M.D.; Keenan, S.; Curtis, M.; Cassidy, M.; Byrne, G.; Destrade, M. Measuring knife stab penetration into skin simulant using a novel biaxial tension device. *Forensic Sci. Int.* **2008**, *177*, 52–65. [CrossRef]
17. Sharkey, E.J.; Cassidy, M.; Brady, J.; Gilchrist, M.D.; NicDaeid, N. Investigation of the force associated with the formation of lacerations and skull fractures. *Int. J. Leg. Med.* **2012**, *126*, 835–844. [CrossRef]
18. Hainsworth, S.V.; Delaney, R.J.; Rutty, G.N. How sharp is sharp? Towards quantification of the sharpness and penetration ability of kitchen knives used in stabbings. *Int. J. Leg. Med.* **2008**, *122*, 281–291. [CrossRef]
19. O'Callaghan, P.T.; Jones, M.D.; James, D.S.; Leadbeatter, S.; Holt, C.A.; Nokes, L.D. Dynamics of stab wounds: Force required for penetration of various cadaveric human tissues. *Forensic Sci. Int.* **1999**, *104*, 173–178. [CrossRef]
20. Chadwick, E.K.; Nicol, A.C.; Lane, J.V.; Gray, T.G. Biomechanics of knife stab attacks. *Forensic Sci. Int.* **1999**, *105*, 35–44. [CrossRef]
21. Whittle, K.; Kieser, J.; Ichim, I.; Swain, M.; Waddell, N.; Livingstone, V.; Taylor, M. The biomechanical modelling of non-ballistic skin wounding: Blunt-force injury. *Forensic Sci. Med. Pathol.* **2008**, *4*, 33–39. [CrossRef] [PubMed]
22. Shergold, O.A.; Fleck, N.A. Experimental investigation into the deep penetration of soft solids by sharp and blunt punches, with application to the piercing of skin. *J. Biomech. Eng.* **2005**, *127*, 838–848. [CrossRef]
23. Parmar, K.; Hainsworth, S.V.; Rutty, G.N. Quantification of forces required for stabbing with screwdrivers and other blunter instruments. *Int. J. Leg. Med.* **2012**, *126*, 43–53. [CrossRef]
24. Ankersen, J.; Birkbeck, A.; Thomson, R.; Vanezis, P. The effect of knife blade profile on penetration force in flesh simulants. *Technol. Law Insur.* **1998**, *3*, 125–128. [CrossRef]
25. Harding, B.; Risdon, R.A.; Krous, H.F. Shaken baby syndrome. *Br. Med J.* **2004**, *328*, 720–721. [CrossRef]
26. Mole, C.G.; Heyns, M.; Cloete, T. How hard is hard enough? An investigation of the force associated with lateral blunt force trauma to the porcine cranium. *Leg. Med.* **2015**, *17*, 1–8. [CrossRef]
27. Delannoy, Y.; Becart, A.; Colard, T.; Delille, R.; Tournel, G.; Hedouin, V.; Gosset, D. Skull wounds linked with blunt trauma (hammer example). A report of two depressed skull fractures—Elements of biomechanical explanation. *Leg. Med. (Tokyo Jpn.)* **2012**, *14*, 258–262. [CrossRef]
28. Sulaiman, N.A.; Osman, K.; Hamzah, N.H.; Amir, S.P. Blunt force trauma to skull with various instruments. *Malays. J. Pathol.* **2014**, *36*, 33–39. [PubMed]
29. Zhou, Z.; Jiang, B.; Cao, L.; Zhu, F.; Mao, H.; Yang, K.H. Numerical simulations of the 10-year-old head response in drop impacts and compression tests. *Comput. Methods Programs Biomed.* **2016**, *131*, 13–25. [CrossRef]
30. Roschger, P.; Paschalis, E.P.; Fratzl, P.; Klaushofer, K. Bone mineralization density distribution in health and disease. *Bone* **2008**, *42*, 456–466. [CrossRef]
31. Lynnerup, N. Cranial thickness in relation to age, sex and general body build in a Danish forensic sample. *Forensic Sci. Int.* **2001**, *117*, 45–51. [CrossRef]
32. De Boer, H.H.H.; Van der Merwe, A.E.L.; Soerdjbalie-Maikoe, V.V. Human cranial vault thickness in a contemporary sample of 1097 autopsy cases: Relation to body weight, stature, age, sex and ancestry. *Int. J. Leg. Med.* **2016**, *130*, 1371–1377. [CrossRef] [PubMed]
33. Rowbotham, S.K.; Mole, C.G.; Tieppo, D.; Blaszkowska, M.; Cordner, S.; Blau, S. Thickness of the human neurocranium: Reference measurements to strengthen interpretations of degree of force in cases of blunt head injury. *Int. J. Leg. Med.* **2021**. Under review.
34. Robertson, B.; Vignaux, G.A.; Berger, C.E.H. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*; John Wiley & Sons: London, UK, 2016.
35. Joyce, J. Bayes' theorem. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2003.

36. Evett, I.W. Bayesian Inference and Forensic Science: Problems and Perspectives. *Statistician* **1987**, *36*, 99–105. [CrossRef]
37. Evett, I.W. The logical foundations of forensic science: Towards reliable knowledge. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2015**, *B 370*, 20140263. [CrossRef]
38. Banks, D.; Tackett, M. Bayesian methods and forensic inference. In *Handbook of Forensic Statistics*; Banks, D., Kafadar, K., Kaye, D.H., Tackett, M., Eds.; CRC Press: Boca Raton, FL, USA, 2021; pp. 73–90.
39. Gee, D.J. Reaching conclusions in forensic pathology. *Med. Sci. Law* **1995**, *35*, 12–16. [CrossRef]
40. De Boer, H.H.; Fronczek, J.; Berger, C.E.H.; Sjerps, M. The logic of forensic pathology opinion. *Int. J. Leg. Med.* **2021**, in press.
41. Steadman, D.W.; Adams, B.J.; Konigsberg, L.W. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.* **2006**, *131*, 15–26. [CrossRef]
42. Langley-Shirley, N.; Jantz, R.L. A Bayesian approach to age estimation in modern Americans from the clavicle. *J. Forensic Sci.* **2010**, *55*, 571–583. [CrossRef] [PubMed]
43. Andersson, M.G.; Sundström, A.; Lindström, S. Bayesian networks for evaluation of evidence from forensic entomology. *Biosecur. Bioterror. Biodef. Strategy Pract. Sci.* **2013**, *11*, S64–S77. [CrossRef]
44. Champod, C.; Meuwly, D. The inference of identity in forensic speaker recognition. *Speech Com.* **2000**, *31*, 193–203. [CrossRef]
45. Freeman, M.D.; Kohles, S.S. Applications and limitations of Forensic Biomechanics: A Bayesian perspective. *J. Forensic Leg. Med.* **2010**, *17*, 67–77. [CrossRef] [PubMed]
46. Andersson, M.G.; Ceciliason, A.-S.; Sandler, H.; Mostad, P. Application of the Bayesian framework for forensic interpretation to casework involving postmortem interval estimates of decomposed human remains. *Forensic Sci. Int.* **2019**, *301*, 402–414. [CrossRef]
47. Cave, R.; DiMaio, V.J.; Molina, D.K. Homicide or suicide? Gunshot wound interpretation: A Bayesian approach. *Am. J. Forensic Med. Pathol.* **2014**, *35*, 118–123. [CrossRef]
48. Christensen, A.M.; Passalacaqua, N.V.; Bartelink, E.J. Personal identification. In *Forensic Anthropology: Current Methods and Practice*, 2nd ed.; Christensen, A.M., Passalacaqua, N.V., Bartelink, E.J., Eds.; Academic Press: London, UK, 2019; pp. 443–468.
49. Berger, C.E.H.; van Wijk, M.; de Boer, H.H. Bayesian inference in personal identification. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, S., Cattaneo, C., Eds.; Academic Press: London, UK, 2020; pp. 301–312.
50. Bruijning-van Dongen, C.J.; Slooten, K.; Burgers, W.; Wiegerinck, W. Bayesian networks for victim identification on the basis of DNA profiles. *Forensic Sci. Int. Genet. Suppl. Ser.* **2009**, *2*, 466–468. [CrossRef]
51. de Boer, H.H.; Blau, S.; Delabarde, T.; Hackman, L. The role of forensic anthropology in disaster victim identification (DVI): Recent developments and future prospects. *Forensic Sci. Res.* **2018**, *4*, 1–14. [CrossRef]
52. Caridi, I.; Alvarez, E.E.; Somigliana, C.; Puerto, M.S. Using already-solved cases of a mass disaster event for prioritizing the search among remaining victims: A Bayesian approach. *Sci. Rep.* **2020**, *10*, 5026. [CrossRef] [PubMed]
53. Budowle, B.; Ge, J.; Chakraborty, R.; Gill-King, H. Use of prior odds for missing persons identifications. *Investig. Genet.* **2011**, *2*, 15. [CrossRef]
54. Anonymous. *ENFSI Guideline for Evaluative Reporting in Forensic Science*; European Network of Forensic Science Institutes (ENFSI): Wiesbaden, Germany, 2015.
55. de Boer, H.H.; van Wijk, M.; Berger, C.E.H. Communicating evidence with a focus on the use of Bayes' theorem. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, S., Cattaneo, C., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 331–340.
56. Zephro, L.; Galloway, A. The biomechanics of fracture production. In *Broken Bones: Anthropological Analysis of Blunt Force Trauma*; Wedel, V.L., Galloway, A., Eds.; Charles C Thomas: Springfield, IL, USA, 2014.
57. Rowbotham, S.K.; Blau, S.; Hislop-Jambrich, J.; Francis, V. Skeletal trauma resulting from fatal low ($\leq$ 3 m) free falls. An analysis of fracture patterns and morphologies. *J. Forensic Sci.* **2017**, *63*, 1010–1020. [CrossRef] [PubMed]
58. Fenton, T.W.; deJong, J.L.; Haut, R.C. Punched with a fist: The etiology of a fatal depressed cranial fracture. *J. Forensic Sci.* **2003**, *48*, 1–5. [CrossRef]

biology

MDPI

Article

A Blood–Bone–Tooth Model for Age Prediction in Forensic Contexts

Helena Correia Dias [1,2,3,*], Licínio Manco [1], Francisco Corte Real [3,4] and Eugénia Cunha [2,3]

[1] Research Centre for Anthropology and Health (CIAS), Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; lmanco@antrop.uc.pt
[2] Centre for Functional Ecology (CEF), Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; eugenia.m.cunha@inmlcf.mj.pt
[3] National Institute of Legal Medicine and Forensic Sciences, 3000-548 Coimbra, Portugal; francisco.m.cortereal@inmlcf.mj.pt
[4] Faculty of Medicine, University of Coimbra, 3000-370 Coimbra, Portugal
* Correspondence: helenacorreiadias30@gmail.com; Tel.: +351-239240700; Fax: +351-239855211

Simple Summary: DNA methylation age estimation is one of the hottest topics in forensic field nowadays. Age estimation can be improved under a multidisciplinary approach, the role of a forensic anthropologist and forensic epigeneticist being crucial in the establishment of new basis for age estimation. The development of epigenetic models for bones and tooth samples is crucial in this way. Moreover, developing models for age estimation using several samples can be a useful tool in forensics. In this study, we built two multi-tissue models for age estimation, combining blood, bones and tooth samples and using two different methodologies. Through the Sanger sequencing methodology, we built a model with seven age-correlated markers and a mean absolute deviation between predicted and chronological ages of 6.06 years. Using the SNaPshot assay, a model with three markers has been developed revealing a mean absolute deviation between predicted and chronological ages of 6.49 years. Our results showed the usefulness of DNA methylation age estimation in forensic contexts and brought new insights into the development of multi-tissue models applied to blood, bones and teeth. In the future, we expected that these procedures can be applied to the Medico-Legal facilities to use DNA methylation in routine practice for age estimation.

Abstract: The development of age prediction models (APMs) focusing on DNA methylation (DNAm) levels has revolutionized the forensic age estimation field. Meanwhile, the predictive ability of multi-tissue models with similar high accuracy needs to be explored. This study aimed to build multi-tissue APMs combining blood, bones and tooth samples, herein named blood–bone–tooth-APM (BBT-APM), using two different methodologies. A total of 185 and 168 bisulfite-converted DNA samples previously addressed by Sanger sequencing and SNaPshot methodologies, respectively, were considered for this study. The relationship between DNAm and age was assessed using simple and multiple linear regression models. Through the Sanger sequencing methodology, we built a BBT-APM with seven CpGs in genes *ELOVL2*, *EDARADD*, *PDE4C*, *FHL2* and *C1orf132*, allowing us to obtain a Mean Absolute Deviation (MAD) between chronological and predicted ages of 6.06 years, explaining 87.8% of the variation in age. Using the SNaPshot assay, we developed a BBT-APM with three CpGs at *ELOVL2*, *KLF14* and *C1orf132* genes with a MAD of 6.49 years, explaining 84.7% of the variation in age. Our results showed the usefulness of DNAm age in forensic contexts and brought new insights into the development of multi-tissue APMs applied to blood, bone and teeth.

Keywords: DNA methylation (DNAm); epigenetic age estimation; multi-tissue age prediction models (APMs); Sanger sequencing; SNaPshot

Citation: Correia Dias, H.; Manco, L.; Corte Real, F.; Cunha, E. A Blood–Bone–Tooth Model for Age Prediction in Forensic Contexts. *Biology* **2021**, *10*, 1312. https://doi.org/10.3390/biology10121312

Academic Editors: Ann H. Ross and Andrés Moya

Received: 29 October 2021
Accepted: 7 December 2021
Published: 10 December 2021

1. Introduction

Age estimation is one of the most important issues in forensic contexts. Among the parameters of the biological profile, the estimate of adult's age at death has always

been problematic in forensic anthropology since skeletal aging continues to be largely unknown, and all the available methods continue to fail in the approximation to the real age. In other words, there is a discrepancy between biological and chronological ages; the older, the worse. Despite significant research that has been conducted to face problems of adults' age at death, there is not a model of age prediction that can be considered very satisfactory. In particular, aging the elderly is lacking age indicators that can discriminate among individuals of seventy, eighty and ninety. Apart from that, the methods that can be applied always depend both on the state of completeness and preservation of the human remains. In forensic anthropology practice, there are many situations where the targeted age indicators are missing and where alternatives are needed. That is the case of some burned remains, dismembered bodies and incomplete bodies, among others. On the other hand, in the case of a fresh body of an unidentified victim, where physiognomic traits are no longer available and with no suspicion of identity, age is always a needed parameter. In those situations, an alternative is also required. Although imaging methods could be a good alternative, we here argue that the genetic approach by means of DNA methylation (DNAm) is also a good choice.

DNAm analysis for age estimation has emerged in the forensic field in recent years. Several age-related markers have been investigated in different tissues, including blood, saliva, buccal swabs, sperm, teeth and bones, allowing the development of tissue-specific age prediction models (APMs) with high accuracy [1]. The development of multi-tissue APMs brought many advantages in forensics, since they can be applied to several contexts with different types of samples. However, the discovery of universal biomarkers of age applied simultaneously to many tissue types can be a challenge, since it has been observed that only a few markers can work well as multi-tissue age predictive markers [2].

To our knowledge, only three reports addressed multi-tissue DNAm age prediction in human individuals. Horvath [3] assessed methylation information of 353 CpGs, developing a highly accurate multi-tissue age predictive model showing a strong correlation between predicted and chronological ages (R = 0.97), and revealing a median absolute difference between chronological and predicted ages of 2.9 years (training set) and 3.6 years (test set). The high accuracy can be explained by the larger number of CpGs included in the model. However, a high number of age markers can also bring a challenge for forensic casework application. Moreover, in the Horvath model a larger error (around 10 years) was observed in several tissues suggesting that the best markers for one tissue may not be the best for another. Using published databases, Alsaleh et al. [4] identified a small set of 10 CpG sites and built a multi-tissue model for blood, semen, saliva, menstrual blood and vaginal secretions with a Mean Absolute Deviation (MAD) from chronological age of 3.8 years. Jung et al. [2] developed a multi-tissue APM applied to blood, buccal swabs and saliva with DNAm captured by a SNaPshot assay using five CpGs located at *ELOVL2, FHL2, C1orf132, KLF14* and *TRIM59* genes. The multi-tissue model showed high accuracy with a MAD from chronological age of 3.553 years. This MAD value was similar to that reported in the same study when developing tissue-specific APMs (MAD = 3.17 years in blood; MAD = 3.82 years in buccal swabs; MAD = 3.29 years in saliva). In addition, Jung and colleagues [2] have observed that the *FHL2* gene is more tissue-specific, revealing strong positive age correlation values in saliva and blood, and a weak age correlation in buccal swabs. They observed also that *ELOVL2* and *TRIM59* seem to work as better multi-tissue markers than *FHL2, C1orf132* or *KLF14*.

Our group previously assessed the methylation information of age-correlated CpG sites in genes *ELOVL2*, *FHL2*, *EDARADD*, *PDE4C*, *C1orf132*, *TRIM59* and *KLF14*, captured by Sanger sequencing and SNaPshot methodologies [5–8]. Several tissue-specific APMs were developed, including for blood [5–7], teeth [8] and bones [8]. Considering the scarcity of multi-tissue APMs developed until now, the present study aimed to re-examine the obtained DNAm levels for these highly age-correlated genes combining the previously addressed tissues to test for a multi-tissue blood–bone–tooth age prediction model (BBT-APM).

2. Materials and Methods

2.1. Population Sample

A total of 185 samples (76 females, 109 males; aged 1–94 years old) from living and deceased individuals from blood, bones and teeth previously addressed for DNAm levels by Sanger sequencing in genes *ELOVL2* (9 CpGs), *EDARADD* (4 CpGs), *FHL2* (12 CpGs), *PDE4C* (12 CpGs) and *C1orf132* (6 CpGs) [5,6,8], and 168 samples (67 females, 101 males; 1–94 aged years old) from living and deceased individuals previously analyzed using a SNaPshot assay for 5 specific CpG sites in genes *ELOVL2, FHL2, KLF14*, *C1orf132* and *TRIM59* [7,8], were considered for this study. The same samples were addressed in both methodologies; however, some samples failed PCR amplification and were excluded from further analysis, which explains the difference in number between the two methods. The age distribution of each training set was shown in Table S1.

Peripheral blood samples from healthy living individuals of Portuguese ancestry were collected from users of Biobanco—Hospital Pediátrico de Coimbra and other hospitals; blood samples from deceased individuals were collected during routine autopsies, after consulting RENNDA (Registo Nacional de Não Dadores) in Serviço de Patologia Forense da Delegação do Centro do Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF) and from Bodies Donated to Science (BDS), before the embalming method in Departamento de Anatomia da Faculdade de Medicina da Universidade do Porto (FMUP). Fresh bone samples (rib) were collected, after consulting RENNDA, during autopsy in Serviço de Patologia Forense das Delegações do Centro e Sul do INMLCF. Tooth samples (molars) from living individuals were collected in dentist offices, after written informed consent, and tooth samples from deceased individuals (molars) were collected from BDS in Departamento de Anatomia da FMUP. We excluded individuals with known diseases or other clinical conditions that could influence DNAm levels. All blood and bone samples from dead bodies were collected within five days after death.

The herein developed multi-tissue APM using Sanger sequencing includes: 65 blood samples from healthy individuals (42 females, 23 males; aged 1–94 years old), 68 blood samples from deceased individuals (15 females, 53 males; aged 24–91 years old), 23 tooth samples (15 females, 8 males; aged 26–88 years old) and 29 bone samples (4 females, 25 males; aged 26–81 years old). For the multi-tissue APM developed by SNaPshot, 55 blood samples from healthy individuals (34 females, 21 males; aged 1–94 years old), 59 blood samples from deceased individuals (13 females, 46 males; aged 24–91 years old), 23 tooth samples (15 females, 8 males; aged 26–88 years old) and 31 bone samples (5 females, 26 males; aged 26–81 years old) were considered.

The study protocol was approved by the ethical Committee of Faculdade de Medicina da Universidade de Coimbra (n° 038-CE-2017). For living individuals, written informed consent was previously obtained from adult participants and from children's parents under the age of 18 years.

2.2. Sanger Sequencing of C1orf132 in Blood Samples from Living Individuals

As the *C1orf132* gene was not previously addressed in blood samples from living individuals using the Sanger sequencing methodology, the genomic DNA extracted from blood samples of living individuals using the *QIAamp DNA Mini Kit* (Qiagen, Hilden, Germany) was bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA), and submitted to polymerase chain reaction (PCR) amplification using the Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany) for a selected region of *C1orf132*, as previously described [5]. Sequencing was performed in the ABI 3130 sequencer (Applied Biosystems, Foster City) with Big-Dye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems), using primers and conditions previously described [5].

2.3. Statistical Analyses

Statistical analyses were performed using IBM SPSS statistics software for Windows, version 24.0 (IBM Corporation, Armonk, NY, USA). Linear regression models were used

to analyze the relationships between DNAm levels at CpG sites and chronological age. The simple linear regression coefficients from the highest age-correlated CpGs from each gene for Sanger sequencing data, and from each age-correlated CpG site addressed by SNaPshot, were used to predict the age of individuals in the combined set of blood, bone and tooth samples. For both methodologies, all the statistically significant age-correlated CpG sites were combined for analysis using the stepwise regression approach for selection of the relevant variables to be included in a multi-locus BBT-APM. We calculated the Spearman correlation value, the mean absolute deviation (MAD) and the root mean square error (RMSE) between chronological and predicted ages for the combined training set of samples in both methodologies. For both the training sets, each obtained MAD value was interpreted as either correct or incorrect using a cutoff value according to the standard error (SE) of the estimate calculated for each APM.

In addition, the MAD values were calculated for subsets of four distinct age categories (<30 years, 31–55 years, 56–79 years, >80 years) for each training set used in Sanger sequencing and SNaPshot methodologies.

Validation of the BBT-APMs was performed by 3-fold cross-validation that consists of randomly removing a set of samples from the training set and to develop three independent multiple linear regressions on the remaining samples. Subsequently, each model is used to predict the age of the removed samples assigned as validation sets. An additional validation was performed by splitting the complete data set into two subsets (training and validation sets) and independent regression was calculated for the training set and applied to the validation set. All the independent linear regression equations developed for validation purposes included the same CpG sites that have been selected for development of the final multi-tissue APM for each methodology.

3. Results

3.1. Multi-Tissue BBT-APM using Sanger Sequencing

DNAm levels of 43 CpGs located at *ELOVL2* (9 CpGs), *EDARADD* (4 CpGs), *FHL2* (12 CpGs), *PDE4C* (12 CpGs) and *C1orf132* (6 CpGs) were assessed in a combined training set of 185 samples, including blood, teeth and bones from Portuguese individuals (76 females, 109 males; aged 1–94 years) using the bisulfite PCR sequencing methodology. The simple linear regression analysis showed that the strongest age-correlated site in each gene was: *ELOVL2* CpG6 (R = 0.759, *p*-value = 6.87×10^{-36}), explaining 57.3% of the variation in age; *FHL2* CpG1 (R = 0.692, *p*-value = 1.11×10^{-27}), explaining 47.6% of the variation in age; *EDARADD* CpG3 (R = -0.682, *p*-value = 1.21×10^{-26}), explaining 46.2% of the variation in age; *C1orf132* CpG1 (R = -0.654, *p*-value = 5.67×10^{-24}), explaining 42.5% of the variation in age and *PDE4C* CpG2 (R = 0.613, *p*-value = 1.79×10^{-20}), explaining 37.2% of the variation in age (Table 1 and Supplementary Table S2). A clear positive age correlation was observed for *ELOVL2* CpG6, *PDE4C* CpG2 and *FHL2* CpG1 markers, and a clear negative age correlation was observed for *EDARADD* CpG3 and *C1orf132* CpG1 markers (Supplementary Figure S1). The predicted age of individuals was calculated using the simple linear regression coefficients for the individual strongest age-associated markers allowing us to obtain MAD values of 12.01 years for *ELOVL2* CpG6, 13.23 years for *C1orf132* CpG1, 13.52 years for *EDARADD* CpG3, 13.16 years for *FHL2* CpG1 and 13.58 years for *PDE4C* CpG2 (Table 1).

Simultaneously testing the 35 significant age-associated CpGs from *ELOVL2* (nine CpGs), *EDARADD* (three CpGs), *FHL2* (nine CpGs), *PDE4C* (eight CpGs) and *C1orf132* (six CpGs) using stepwise regression analysis allowed us to select a multi-locus APM combining seven CpGs (*EDARADD* CpG3, *FHL2* CpG5, *FHL2* CpG11, *ELOVL2* CpG5, *PDE4C* CpG5, *PDE4C* CpG9, *C1orf132* CpG3). The multiple regression analysis combining these CpGs enabled an age correlation (R) value of 0.940 (*p*-value = 7.36×10^{-79}), explaining 87.8% of the variation in age (corrected R^2 = 0.878) (Table 1). The formula to predict age of individuals built with the multiple linear regression coefficients (Supplementary Table S3) was as follows: $26.852 - 24.767 \times$ DNAm level *EDARADD* CpG3 $+ 68.537 \times$ DNAm

level *FHL2* CpG5 − 51.319 × DNAm level *FHL2* CpG11 + 57.461 × DNAm level *ELOVL2* CpG5 + 41.449 × DNAm level *PDE4C* CpG5 − 66.397 × DNAm level *PDE4C* CpG9 − 27.418 × DNAm level *C1orf132* CpG3. The correlation between predicted and chronological ages was 0.915 (Spearman correlation coefficient) with a MAD from chronological age of 6.06 years (RMSE = 7.60) (Figure 1). Correct predictions were 73%, assuming that chronological and predicted ages match around eight years, according to the standard error of estimate calculated for the final APM (SE = 7.86).

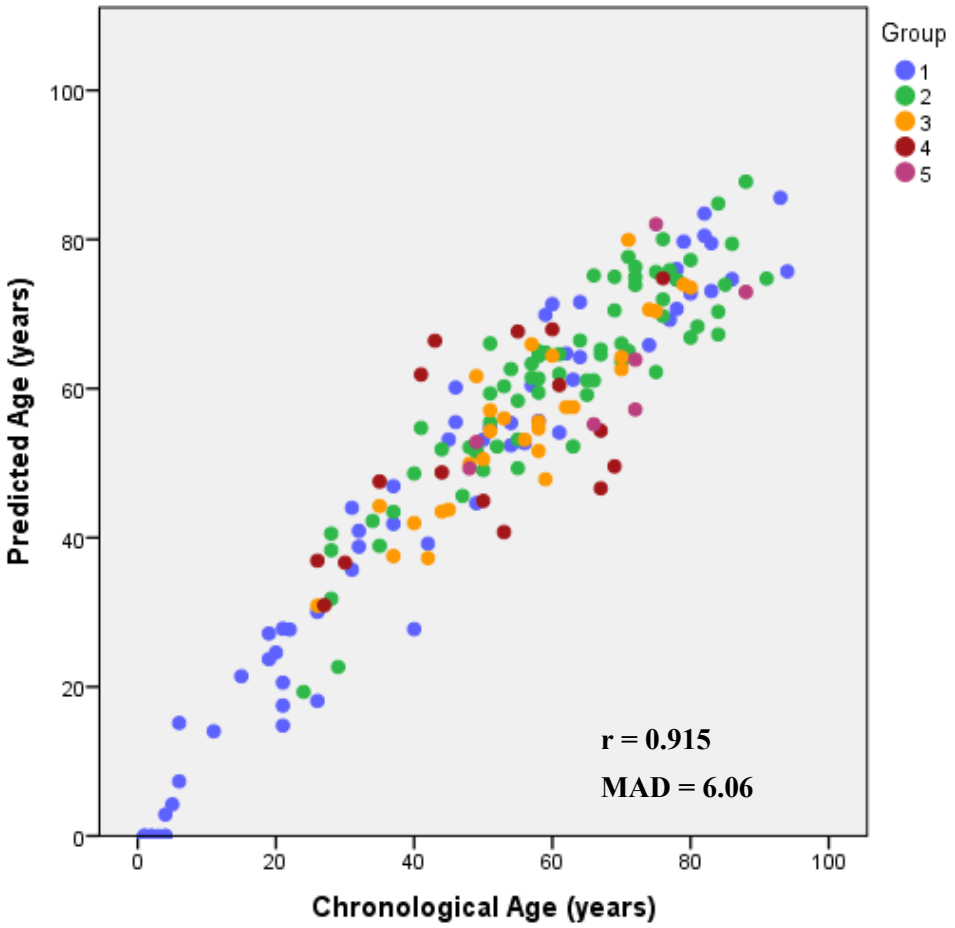


Figure 1. Predicted age versus chronological age using the multi-locus multi-tissue APM developed for *ELOVL2*, *FHL2*, *EDARADD*, *PDE4C* and *C1orf132* genes including blood samples from living individuals (1), blood samples from deceased individuals (2), bone samples (3), tooth samples from living individuals (4) and tooth samples from deceased individuals (5). The corresponding Spearman correlation coefficients (r) are depicted inside each plot.

The accuracy of the developed BBT-APM was tested through a threefold cross validation in the training set of 185 samples showing a MAD of 6.27 years (RMSE = 6.27) (mean value obtained for the three test sets). This value was very close to the MAD of 6.06 (RMSE = 7.60) obtained in the whole training set. The validation by splitting the overall training set into two sets of 117 and 68 samples (training and validation sets) allowed us to obtain an independent MAD value for the training set of 6.09 years (RMSE = 7.55); applying the model on the validation set, a MAD of 6.08 years (RMSE = 7.64) was obtained. Both independent MAD values were very close to the MAD of 6.06 (RMSE = 7.60) obtained in the whole training set.

3.2. Multi-Tissue BBT-APM Using SNaPshot Methodology

DNAm levels at five CpG sites from the *ELOVL2*, *FHL2*, *KLF14*, *C1orf132* and *TRIM59* genes obtained through a SNaPshot assay were simultaneously addressed in a combined

set of 168 samples, including blood, bones and teeth (67 females, 101 males; 1–94 aged years old). DNAm levels of *ELOVL2*, *FHL2*, *KLF14* and *TRIM59* genes revealed a positive correlation with age, and DNAm levels of *C1orf132* showed a negative correlation (Supplementary Figure S2). Testing the individual DNAm association with chronological age for the five CpG sites, the strongest correlation was observed for *ELOVL2* (R = 0.772, *p*-value = 1.54×10^{-34}), explaining 59.4% of the variation in age, followed by *C1orf132* (R = −0.693, *p*-value = 2.49×10^{-25}), explaining 47.7% of the variation in age, *FHL2* (R = 0.686, *p*-value = 1.36×10^{-24}), explaining 46.8% of the variation in age, *KLF14* (R = 0.677, *p*-value = 6.57×10^{-24}), explaining 45.6% of the variation in age and *TRIM59* (R = 0.584, *p*-value = 1.17×10^{-16}), explaining 33.7% of the variation in age (Table 2). The simple APMs for each CpG site allowed us to obtain MAD values from a chronological age of 10.95 years for *ELOVL2*, 12.10 years for *C1orf132*, 12.63 years for *FHL2*, 12.74 years for *KLF14* and 13.64 years for *TRIM59* (Table 2).

Applying the stepwise regression approach to the five CpG sites, only the CpGs located at *ELOVL2, KLF14* and *C1orf132* genes were selected for the development of a final multi-locus APM. The three selected CpGs revealed in the multiple regression analysis a very strong correlation with age, R = 0.922 (*p*-value = 3.14×10^{-67}), explaining 84.7% of the variation in age (corrected R^2 = 0.847) (Table 2). Predicted age through the multivariate regression coefficients was as follows (Supplementary Table S4): 29.220 + 96.850 × DNAm level *ELOVL2* + 208.747 × DNAm level *KLF14* − 33.437 × DNAm level *C1orf132.* This BBT-APM allowed us to obtain a MAD from chronological age of 6.49 years (RMSE = 8.42) (Table 2). Correct predictions were 73.8% considering the cutoff of 9 years, according to the standard error of estimate calculated for the final APM (SE = 8.53). The obtained Spearman correlation value between predicted and chronological ages was 0.893 (Figure 2).

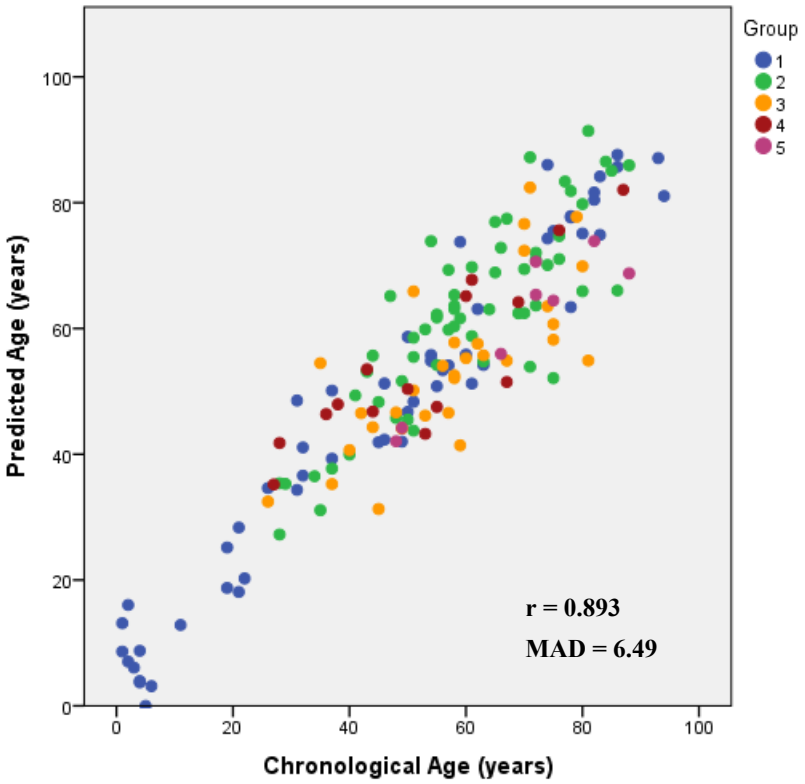


Figure 2. Predicted age versus chronological age using the multi-tissue APM developed for *ELOVL2*, *C1orf132* and *KLF14* genes including blood samples from living individuals (1), blood samples from deceased individuals (2), bone samples (3), tooth samples from living individuals (4) and tooth samples from deceased individuals (5). The corresponding Spearman correlation coefficients (r) are depicted inside each plot.

Table 1. Simple and multiple linear regression statistics of the age predictors in *ELOVL2, FHL2, EDARADD*, *PDE4C* and *C1orf132* genes to test for association between the DNAm levels and chronological age using Sanger sequencing methodology.

Locus	CpG Site	Location	Multi-Tissue: Type of Samples Included	N	R	R^2	Corrected R^2	SE	*p*-Value	MAD
Simple linear regression										
ELOVL2	CpG6	Chr6:11044644	Blood * + Bones + Teeth	185	0.759	0.576	0.573	14.70	6.87×10^{-36}	12.01
FHL2	CpG1	Chr2:105399282	Blood * + Bones + Teeth	185	0.692	0.479	0.476	16.29	1.11×10^{-27}	13.16
EDARADD	CpG3	Chr1:236394382	Blood * + Bones + Teeth	185	−0.682	0.465	0.462	16.51	1.21×10^{-26}	13.52
C1orf132	CpG1	Chr1:207823681	Blood * + Bones + Teeth	185	−0.654	0.428	0.425	17.07	5.67×10^{-24}	13.23
PDE4C	CpG2	Chr19:18233133	Blood * + Bones + Teeth	185	0.613	0.376	0.372	17.83	1.79×10^{-20}	13.58
Multiple linear regression										
APM (*EDARADD* CpG3, *FHL2* CpG5, *FHL2* CpG11, *ELOVL2* CpG5, *PDE4C* CpG5, *PDE4C* CpG9, *C1orf132* CpG3)			Blood * + Bones + Teeth	185	0.940	0.883	0.878	7.86	7.36×10^{-79}	6.06

Abbreviations: N, number of samples; R, correlation coefficient; SE, standard error; MAD, mean absolute deviation (years) between chronological and predicted ages. Genomic positions were based on the GRCh38/hg38 assembly. * Blood samples from living and deceased individuals.

Table 2. Simple and multiple linear regression statistics at the five CpGs of the *ELOVL2, FHL2, KLF14, TRIM59* and *C1orf132* genes to test for association between the DNAm levels and chronological age using SNaPshot assay.

Locus	Location	Multi-Tissue: Type of Samples Included	N	R	R^2	Corrected R^2	SE	*p*-Value	MAD
Simple linear regression									
ELOVL2	Chr6:11044628	Blood * + Bones + Teeth	168	0.772	0.597	0.594	13.896	1.54×10^{-34}	10.95
FHL2	Chr2:105399282	Blood * + Bones + Teeth	168	0.686	0.471	0.468	15.885	1.36×10^{-24}	12.63
KLF14	Chr7:130734355	Blood * + Bones + Teeth	168	0.677	0.459	0.456	16.091	6.57×10^{-24}	12.74
C1orf132	Chr1:207823681	Blood * + Bones + Teeth	168	−0.693	0.480	0.477	15.779	2.49×10^{-25}	12.10
TRIM59	Chr3:160450189	Blood * + Bones + Teeth	168	0.584	0.341	0.337	17.780	1.17×10^{-16}	13.64
Multiple linear regression									
APM (*ELOVL2, KLF14* and *C1orf132*)		Blood * + Bones + Teeth	168	0.922	0.850	0.847	8.53	3.14×10^{-67}	6.49

Abbreviations: N, number of samples; R, correlation coefficient; SE, standard error; MAD, mean absolute deviation (years) between chronological and predicted ages. Genomic positions were based on the GRCh38/hg38 assembly. * Blood samples from living and deceased individuals.

The model accuracy of the final APM with DNAm levels of *ELOVL2*, *KLF14* and *C1orf132* markers was evaluated through a threefold cross validation in the training set of 168 samples, producing a MAD (mean value obtained for the three test sets) of 6.73 years (RMSE = 6.75). This value was very close to the MAD of 6.49 (RMSE = 8.42) obtained in the whole training set. The validation by splitting the overall training set into two sets of 113 and 55 samples (training and validation sets) allowed us to obtain an independent MAD value for the training set of 6.06 years (RMSE = 7.81). Applying the model on the validation set, a MAD of 7.45 years (RMSE = 9.60) was obtained.

3.3. Differences between Predicted and Chronological Ages with an Increase in Age

Evaluating the model performance obtained with the two developed multi-tissue BBT-APMs according to different age ranges (Table 3), we observed an increase in the MAD values between predicted and chronological ages with the increase in age of individuals. For both Sanger sequencing and SNaPshot methodologies, the value of MAD was the largest for the age group >80 years and the smallest for age group <30 years (Table 3).

Table 3. Evaluation of mean absolute deviation (MAD) between chronological and predicted ages according to four age-range groups in the training set of blood, bone and tooth samples using both methodologies.

Age Range / Method	Sanger Sequencing		SNaPshot	
	N	MAD (Years)	N	MAD (Years)
<30 years	33	4.73	23	5.51
31–55 years	58	6.37	56	6.23
56–79 years	74	5.67	68	6.74
>80 years	20	8.81	21	7.37

4. Discussion

In the past decade, several specific epigenetic clocks with high accuracy have been developed using many tissue types [9–16]. However, the discovery of DNAm age-related markers with similarly high accuracy across different types of tissues (universal markers) remains a challenging task in the forensic field [17]. Evidence from previous studies shows that each age-correlated marker reveals a specific ability to predict chronological age, as each tissue type can be affected by different intrinsic or environmental factors. Eipel et al. [16] reported that using a specific APM with methylation information of age-correlated markers selected in one tissue-specific type can lead to a decrease in model accuracy in age prediction if applied to a different tissue. This should be related to the tissue-specific differences in epigenetic patterns [18–20]. Thus, a careful selection of age-associated CpGs and the validation of these proposed markers in each tissue type should be the first step for the development of multi-tissue APMs.

In fact, until now, only a few studies have explored the predictive ability of multi-tissue APMs [2–4]. In this study, we re-examined DNAm levels of *ELOVL2, FHL2, PDE4C, EDARADD, C1orf132, TRIM59* and *KLF14* genes, previously captured in different tissue types (blood samples from living and deceased individuals; tooth samples from living and deceased individuals; fresh bone samples collected during autopsies) by Sanger sequencing and SNaPshot methodologies to build multi-tissue APMs. We developed simple linear regression APMs for the best-selected CpG sites from each gene, and multi-locus multi-tissues APMs using the best combination of CpGs selected by the stepwise regression approach.

DNAm levels captured by bisulfite Sanger sequencing allowed the development of a final APM with seven CpGs (*EDARADD* CpG3, *FHL2* CpG5, *FHL2* CpG11, *ELOVL2* CpG5, *PDE4C* CpG5, *PDE4C* CpG9, *C1orf132* CpG3), revealing a very strong age correlation value (R = 0.940), highly significant (p-value = 7.36×10^{-79}) and explaining 87.8% of the variation

in age. The BBT-APM developed with 185 Portuguese individuals (aged 1–94 years old) allows us to predict age with a moderate accuracy showing a MAD from chronological age of 6.06 years.

Regarding methylation information captured by the SNaPshot methodology, the final multi-locus multi-tissue APM combines three CpG sites located at *ELOVL2, KLF14* and *C1orf132* genes. This BBT-APM developed in 168 samples revealed a very strong age correlation value (R = 0.922), with a MAD from chronological age of 6.49 years.

In Table 4, we resume in brief the difference in results obtained with both methodologies.

The multi-tissue APMs developed herein allows prediction of age of the individuals based on evaluation of DNAm levels captured from several types of samples, including blood, bone and teeth. The final models revealed an accuracy (MAD value) of about 6 years, being more accurate than the majority of anthropological approaches applied to adults' age estimation. When comparing the results with the ones retrieved by anthropological methods, it becomes clear that our method has clear benefits in relation to methods such as Suchey–Brooks', where age ranges are particularly large, mainly for old individuals.

Table 4. Comparison between Sanger sequencing and SNaPshot methodologies.

Method	Sanger Sequencing	SNaPshot
CpGs and genes included in the APM	7 CpGs located at 5 genes (*EDARADD* CpG3, *FHL2* CpG5, *FHL2* CpG11, *ELOVL2* CpG5, *PDE4C* CpG5, *PDE4C* CpG9, *C1orf132* CpG3)	3 CpGs located at 3 genes (*ELOVL2, KLF14, C1orf132*)
Age correlation value	0.940	0.922
Variance in age explained	87.8%	84.7%
Accuracy (MAD)	6.06 years	6.49 years
Results	Using the Sanger sequencing methodology, more CpGs and genes were included in the APM, but higher age correlation, higher explained variance in age, and a better accuracy in age prediction (lower MAD value) were obtained.	

Comparing the herein developed multi-tissue APMs with the tissue-specific APMs previously developed by our group, we can observe that through Sanger sequencing, the blood-living APM [6] revealed a MAD of 5.35 years, which is a slightly lower value comparing with the BBT-APM (MAD = 6.06 years). However, for blood samples from deceased individuals [5], the tissue-specific APM revealed a similar accuracy with a MAD of 6.08 years. The tissue-specific APMs developed through the SNaPshot assay for blood samples revealed MAD values of 4.25 and 5.36 years for living and deceased individuals, respectively [7]. However, although these models have a better accuracy than the herein developed BBT-APM using the SNaPshot methodology (MAD = 6.49), they can only be applied to blood samples.

Regarding bones, we have previously obtained through Sanger sequencing and SNaPshot methodologies MAD values of 2.56 and 7.18 years, respectively [8]. Thus, we can observe that for age prediction in bones using Sanger sequencing, it is more advantageous to apply the tissue-specific model compared with the BBT-APM (MAD = 6.06 years). However, using the SNaPshot methodology we obtained a similar prediction accuracy for both the specific bone-APM (MAD = 7.18 years) and the BBT-APM (MAD = 6.49 years). In regards to tooth samples, the tissue-specific models previously developed [8] revealed MAD values of 11.35 years and 7.07 years using Sanger sequencing and SNaPshot methodologies, respectively, which is a lower accuracy in comparison with the BBT-APMs developed in this present study (MAD = 6.06 and 6.49 years, respectively).

Previous reports using DNAm levels for the development of multi-tissues APMs [2–4] showed higher prediction accuracy in age estimation (MAD values of 2.9, 3.55 and 3.8 years). In our study, the obtained higher MAD values (6.06 years in Sanger sequencing and 6.49 years in SNaPshot) can be explained by sample size, population variability or the laboratory methodologies for DNAm assessment. Of note, both developed BBT-APMs

included CpGs from the *ELOVL2* gene revealing the powerful of this age-associated gene for the development of multi-tissue APMs in forensic contexts. It has been shown that *ELOVL2* is a stable epigenetic marker, revealing a high performance as a multi-tissue predictor [2,13,14,21]. This locus has been used as a powerful age-correlated marker in many tissue-specific APMs developed for blood, tooth, bones and buccal swabs, revealing similar patterns of high accuracy in all APMs [2,10–15,22–30]. Moreover, it has been shown that CpGs from the other genes addressed in the present study also revealed higher age correlation values in blood samples [2,5–7,10–12,23,24,26,28–30], bones [8,13,14] and tooth samples [8,15,23,27], being promising markers to be selected for development of universal APMs.

Several aspects should be highlighted for future potential applicability of the herein-developed multi-tissues APMs.

In this study, both BBT-APMs revealed a general decrease in model accuracy (increase in MAD value) with the increase in age, in accordance with previous studies [3,11,12,26,30], revealing that age estimation based on DNAm levels can have a better performance in younger age ranges. Indeed, younger individuals show lower values of MAD reflecting a high accuracy in the APMs, comparing to older ages. This reflects larger differences between biological and chronological ages with the increase in age, related to the accumulation of specific alterations in DNAm patterns of each individual with aging due the stochastic or environmental factors, being accepted as the epigenetic drift contribution [31–33].

The possibility that postmortem changes can alter the methylation status among specific loci should also be hypothesized, and this issue needs future clarification. As reported in previous studies from our group, comparing blood samples from living and deceased individuals [6,7], it is important for forensic casework application to know the healthy status of the sample donor. This is a paramount issue because the most developed APMs until now have been built using samples of living individuals. It has been observed that ancient DNA (aDNA) can suffer postmortem miscoding lesions, as deamination [34,35]. Postmortem deamination is a spontaneously chemical process that occurs due to the hydrolytic deamination of cytosine (C) residues into uracils (U) [34]. If DNA damage in the form of deamination occurs, the expected residues in PCR amplification could be different after bisulfite conversion. Bisulfite conversion is a chemical modification, which mediates the deamination of unmethylated C to U, appearing after PCR amplification as thymine (T), but leaves methylated C (5mC) intact. Therefore, if postmortem cytosine deamination occurs, both unmethylated C and 5mC appear as T after PCR amplification of bisulfite-converted samples, which could disturb the measurement of DNAm levels. As hydroxymethylcytosine (5hmC) is an oxidative product of demethylation of 5mC [36,37], in case of postmortem deamination, the 5hmC concentration can also be affected. Despite this, the stability of 5mC patterns in aDNA has been reported, when preserved aDNA samples were analyzed [38,39]. Moreover, Pedersen et al. [40] assessed to DNAm levels of permafrost hair samples collected from a Paleo-Eskimo with 4000 years old, and predicted age at death. This reveals the reliability on the assessment of DNAm levels to predict age in ancient samples.

An additional important issue for forensic practice is the effect of postmortem interval (PMI) on DNAm levels captured from aged forensic samples of different tissues. Data obtained from such forensic samples should be interpreted with caution due to the very low amount and degradation of the obtained DNA. A previous study developed by Zbieć-Piekarska et al. [24] showed the stability of prediction accuracy using bloodstains that differed in time of storage. The authors evaluated DNA concentrations from bloodstains that had been deposited previously on tissue paper and kept at room temperature during 5, 10 and 15 years, observing a significant decrease in DNA concentration, a decrease in number of positive PCR amplifications and an increase in average degradation index. However, they did not observe an effect on the rate of corrected predictions, reporting that "the prediction success rate seemed not to correlate inversely with increasing time of storage" [24]. Hence, it seems that DNA degradation affects DNA concentration and, con-

sequently, the rate of positive PCR amplifications; however, the accuracy of age prediction is not affected in positive PCR amplification samples.

The major drawback of our study was the limited number of samples, mainly in bones and teeth. We recognize that larger sample sets have greater statistical power and may be more representative of DNAm changes related to different age groups and different types of tissues, leading to the development of more accurate APMs. Another relevant factor that should be considered is the existence of some diseases or clinical conditions or even some life routines such as smoking or drinking, which may interfere with methylation data. For samples of deceased individuals, despite having access to medical reports of each case, information related to possible clinical conditions was unknown in many cases. Lastly, the use of different methodologies for evaluation of DNAm levels across studies can influence the accuracy of APMs. In particular, bisulfite sequencing or SNaPshot methodologies are semi-quantitative methods and thus may not be the optimal tool for precise DNAm analysis.

DNAm analysis is considered a promising method for age estimation in the future. If we question how easy it is to use it and how long it takes to apply it, we argue that in those laboratories supported by genetic facilities provided with the needed equipment, the results can be retrieved in two or three days. In comparison with the more traditional approaches, it takes longer, but in terms of the delivery of the final report, it does not imply any delay. Furthermore, it should be noted that any method that involves DNA analysis turns out to be more expensive, but it also tends to be more reliable. However, it should be emphasized that the development of universal APMs based on DNAm levels is at the beginning of age estimation research and, therefore, the herein proposed BBT-APMs can be a starting point for future research.

5. Conclusions

In conclusion, in this study we re-examined DNAm levels of *ELOVL2, FHL2, PDE4C, EDARADD, C1orf132, TRIM59* and *KLF14* genes previously captured by Sanger sequencing and SNaPshot methodologies across several tissues. Two multi-tissue BBT-APMs were developed using blood, tooth and bone samples from Portuguese individuals. To the best of our knowledge, the two BBT-APMs developed herein for the Portuguese population are the first multi-tissue APMs using bones and teeth. Moreover, despite being very often found in forensic contexts, the development of tissue-specific APMs using bones or teeth is scarce in forensic research. By Sanger sequencing, a moderate accuracy of 6.06 years was obtained in the BBT-APM using seven CpGs from genes *ELOVL2, FHL2, PDE4C, EDARADD* and *C1orf132*. Using the SNaPshot assay, the BBT-APM developed with methylation data from *C1orf132*, *ELOVL2* and *KLF14* genes revealed a MAD from chronological age of 6.49 years. Both methodologies revealed similar accuracy for use in multi-tissue APMs being both simple, rapid, cost-effective and easily available in forensic laboratories. Therefore, both BBT-APMs developed herein can be a promising tool for age estimation in forensic contexts.

This article, a priori, could appear too technical and a little far away from the forensic anthropology reality. However, we argue that a bridge between forensic genetics and forensic anthropology can be achieved, once the needed complicities between the experts involved are well established. In practical terms, what we here advise is an integrated evaluation of the case by the forensic anthropologist, along with the pathologist in charge of the case. If, for instance, the case is a fresh body without any physiognomic traits and where identification is unknown, blood is the best option for DNAm age estimation. If, on the other hand, blood is no longer available due to the state of decomposition of the body, a decision can be made to recover both bone and teeth to perform DNAm studies. What does that imply in practical terms? It means that the result will take 2 or 3 days to be known, that the needed equipment is necessary as well as the adequate kits. While those ones are more expensive than the blood ones, it is a good option in particular when the most adequate skeletal age indicators are damaged or no longer available. Having said

that, we argue that we should strive to implement the procedures here described in the Medico–Legal facilities in order to turn DNAm a routine practice for age estimation.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biology10121312/s1: Figure S1: Correlations between DNAm levels and chronological age in 185 samples including blood samples from living and deceased individuals, bone samples collected from autopsies and teeth from living and deceased individuals, obtained through Sanger sequencing methodology. Figure S2: Correlations between DNAm levels and chronological age in 168 samples, including blood samples from living and deceased individuals, bone samples collected from autopsies and teeth from living and deceased individuals, obtained through SNaPshot methodology. Table S1: Age distribution of the sample sets analyzed by Sanger sequencing and SNaPshot methodologies., Table S2: Univariate linear regression analysis of the 43 CpG sites in *ELOVL2*, *FHL2*, *EDARADD*, *PDE4C* and *C1orf132* loci in 185 samples including blood from living and deceased individuals, teeth from living and deceased individuals and bone collected during autopsies. Table S3: Statistical parameters obtained in a multiple regression model with the seven CpGs in genes *ELOVL2*, *FHL2*, *EDARADD*, *PDE4C* and *C1orf132* selected by stepwise regression approach, in blood, bone and tooth samples. Table S4: Statistical parameters obtained in a multiple regression model with the three CpGs in genes *ELOVL2*, *C1orf132* and *KLF14*, selected by stepwise regression approach, in blood, bone and tooth samples.

Author Contributions: Conceptualization, H.C.D., L.M. and E.C.; data curation, H.C.D.; investigation, H.C.D. and L.M.; methodology, H.C.D. and L.M.; supervision, L.M., F.C.R. and E.C.; writing—original draft, H.C.D.; writing—review and editing, L.M., F.C.R. and E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has funding by Fundação para a Ciência e a Tecnologia (FCT) (UIDB/00283/2020). H.C.D. had a PhD grant from FCT (SFRH/BD/117022/2016).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Faculdade de Medicina da Universidade de Coimbra (n° 038-CE-2017, 26/06/2017) and by the Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank to all the volunteers that participated in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goel, N.; Karira, P.; Garg, V.K. Role of DNA methylation in human age prediction. *Mech. Ageing Dev.* **2017**, *166*, 33–41. [CrossRef]
2. Jung, S.E.; Lim, S.M.; Hong, S.R.; Lee, E.H.; Shin, K.J.; Lee, H.Y. DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Sci. Int. Genet.* **2019**, *38*, 1–8. [CrossRef]
3. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**, *14*, R115. [CrossRef]
4. Alsaleh, H.; McCallum, N.A.; Halligan, D.L.; Haddrill, P.R. A multi-tissue age prediction model based on DNA methylation analysis. *Forensic Sci. Int. Genet. Suppl. Ser.* **2017**, *6*, e62–e64. [CrossRef]
5. Correia Dias, H.; Cordeiro, C.; Corte Real, F.; Cunha, E.; Manco, L. Age estimation based on DNA methylation using blood samples from deceased individuals. *J. Forensic Sci.* **2019**, *65*, 465–470. [CrossRef]
6. Correia Dias, H.; Cunha, E.; Corte Real, F.; Manco, L. Age prediction in living: Forensic epigenetic age estimation based on blood samples. *Leg. Med.* **2020**, *47*, 101763. [CrossRef] [PubMed]
7. Correia Dias, H.; Cordeiro, C.; Pereira, J.; Pinto, C.; Corte Real, F.; Cunha, E.; Manco, L. DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. *Forensic Sci. Int.* **2020**, *311*, 110267. [CrossRef]
8. Correia Dias, H.; Corte Real, F.; Cunha, E.; Manco, L. DNA methylation age estimation from human bone and teeth. *Aust. J. Forensic Sci.* **2021**, Ahead-of-print. 1–14. [CrossRef]
9. Bocklandt, S.; Lin, W.; Sehl, M.E.; Sanchez, F.J.; Sinsheimer, J.S.; Horvath, S.; Vilain, E. Epigenetic predictor of age. *PLoS ONE* **2011**, *6*, e14821. [CrossRef] [PubMed]

10. Garagnani, P.; Bacalini, M.G.; Pirazzini, C.; Gori, D.; Giuliani, C.; Mari, D.; Di Blasio, A.M.D.; Gentilini, D.; Vitale, G.; Collino, S.; et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* **2012**, *11*, 1132–1134. [CrossRef] [PubMed]
11. Bekaert, B.; Kamalandua, A.; Zapico, S.C.; Van de Voorde, W.; Decorte, R. Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* **2015**, *10*, 922–930. [CrossRef]
12. Zbieć-Piekarska, R.; Spólnicka, M.; Kupiec, T.; Parys-Proszek, A.; Makowska, Z.; Pałeczka, A.; Kucharczyk, K.; Płoski, R.; Branicki, W. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci. Int. Genet.* **2015**, *17*, 173–179. [CrossRef]
13. Naue, J.; Sänger, T.; Hoefsloot, H.C.J.; Lutz-Bonengel, S.; Kloosterman, A.D.; Verschure, P.J. Proof of concept study of age-dependent DNA methylation markers across different tissues by massive parallel sequencing. *Forensic Sci. Int. Genet.* **2018**, *36*, 152–159. [CrossRef] [PubMed]
14. Gopalan, S.; Gaige, J.; Henn, B.M. DNA methylation-based forensic age estimation in human bone. *BioRxiv* **2019**, 801647. [CrossRef]
15. Márquez-Ruiz, A.B.; González-Herrera, L.; Luna, J.D.; Valenzuela, A. DNA methylation levels and telomere length in human teeth: Usefulness for age estimation. *Int. J. Leg. Med.* **2020**, *134*, 451–459. [CrossRef]
16. Eipel, M.; Mayer, F.; Arent, T.; Ferreira, M.R.P.; Birkhofer, C.; Costa, I.G.; Ritz-Timme, S.; Wagner, W. Epigenetic age predictions based on buccal swabs are more precise in combination with cell type—Specific DNA methylation signatures. *Aging* **2016**, *8*, 1034–1048. [CrossRef]
17. Koch, M.C.; Wagner, W. Epigenetic-aging-signature to determine age in different tissues. *Aging* **2011**, *3*, 1018. [CrossRef]
18. Illingworth, R.; Kerr, A.; Desousa, D.; Jørgensen, H.; Ellis, P.; Stalker, J.; Jackson, D.; Clee, C.; Plumb, R.; Rogers, J.; et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **2008**, *6*, e22. [CrossRef] [PubMed]
19. Li, Y.; Zhu, J.; Tian, G.; Li, N.; Li, Q.; Ye, M.; Zheng, H.; Yu, J.; Wu, H.; Sun, J.; et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* **2010**, *8*, e1000533. [CrossRef]
20. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; et al. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048. [CrossRef]
21. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sadda, S.; Klotzle, B.; Bibikova, M.; Fan, J.-B.; Gao, Y.; et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell.* **2013**, *49*, 359–367. [CrossRef]
22. Weidner, C.I.; Lin, Q.; Koch, C.M.; Eisele, L.; Beier, F.; Ziegler, P.; Bauerschlag, D.O.; Jöckel, K.H.; Erbel, R.; Mühleisen, T.W.; et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **2014**, *15*, R24. [CrossRef]
23. Bekaert, B.; Kamalandua, A.; Zapico, S.C.; Van de Voorde, W.; Decorte, R. A selective set of DNA-methylation markers for age determination of blood, teeth and buccal samples. *Forensic Sci. Int. Genet. Suppl. Ser.* **2015**, *5*, e144–e145. [CrossRef]
24. Zbieć-Piekarska, R.; Spólnicka, M.; Kupiec, T.; Makowska, Z.; Spas, A.; Parys-Proszek, A.; Kucharczyk, K.; Płoski, R.; Branicki, W. Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. *Forensic Sci. Int. Genet.* **2015**, *14*, 161–167. [CrossRef] [PubMed]
25. Xu, C.; Qu, H.; Wang, G.; Xie, B.; Shi, Y.; Yang, Y.; Zhao, Z.; Hu, L.; Fang, X.; Yan, J.; et al. A novel strategy for forensic age prediction by DNA methylation and support vector regression model. *Nature* **2015**, *5*, 17788. [CrossRef]
26. Hamano, Y.; Manabe, S.; Morimoto, C.; Fujimoto, S.; Ozeki, M.; Tamaki, K. Forensic age prediction for dead or living samples by use of methylation-sensitive high resolution melting. *Leg. Med.* **2016**, *21*, 5–10. [CrossRef] [PubMed]
27. Giuliani, C.; Cilli, E.; Bacalini, M.G.; Pirazzini, C.; Sazzini, M.; Gruppioni, G.; Franceschi, C.; Garagnani, P.; Luiselli, D. Inferring chronological age from DNA methylation patterns of human teeth. *Am. J. Phys. Anthropol.* **2016**, *159*, 585–595. [CrossRef] [PubMed]
28. Freire-Aradas, A.; Phillips, C.; Mosquera-Miguel, A.; Girón-Santamaría, L.; Gómez-Tato, A.; de Cal, M.C.; Álvarez-Dios, J.; Ansede-Bermejo, J.; Torres-Español, M.; Schneider, P.M.; et al. Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Sci. Int. Genet.* **2016**, *24*, 65–74. [CrossRef] [PubMed]
29. Cho, S.; Jung, S.-E.; Hong, S.R.; Lee, E.H.; Lee, J.H.; Lee, S.D.; Lee, H.Y. Independent validation of DNA-based approaches for age prediction in blood. *Forensic Sci. Int. Genet.* **2017**, *29*, 250–256. [CrossRef] [PubMed]
30. Thong, Z.; Chan, X.L.S.; Tan, J.Y.Y.; Loo, E.S.; Syn, C.K.C. Evaluation of DNA methylation-based age prediction on blood. *Forensic Sci. Int. Genet. Suppl. Ser.* **2017**, *6*, e249–e251. [CrossRef]
31. Jones, M.J.; Goodman, S.J.; Kobor, M.S. DNA methylation and healthy human aging. *Aging Cell* **2015**, *14*, 924–932. [CrossRef] [PubMed]
32. Tan, Q.; Heijmans, B.T.; Hjelmborg, J.V.; Soerensen, M.; Christensen, K.; Christiansen, L. Epigenetic drift in the aging genome: A ten-year follow-up in an elderly twin cohort. *Int. J. Epidemiol.* **2016**, *45*, 1146–1158. [CrossRef]
33. Freire-Aradas, A.; Phillips, C.; Lareu, M.V. Forensic individual age estimation with DNA: From initial approaches to methylation tests. *Forensic Sci. Rev.* **2017**, *29*, 121–144. [PubMed]
34. Hofreiter, N.; Jaenicke, V.; Serre, D.; von Haeseler, A.; Paabo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **2001**, *29*, 4793–4799. [CrossRef] [PubMed]

35. Brotherton, P.; Endicott, P.; Sanchez, J.J.; Beaumont, M.; Barnett, R.; Austin, J.; Cooper, A. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* **2007**, *35*, 5717–5728. [CrossRef]
36. Dahl, C.; Grønbæk, K.; Guldberg, P. Advances in DNA methylation: 5-hydroxymethylcytosine revisited. *Clin. Chim. Acta* **2011**, *412*, 831–836. [CrossRef]
37. Shen, L.; Zhang, Y. 5-Hydroxymethylcytosine: Generation, fate, and genomic Distribution. *Curr. Opin. Cell Biol.* **2013**, *25*, 289–296. [CrossRef]
38. Briggs, A.W.; Stenzel, U.; Meyer, M.; Krause, J.; Kircher, M.; Paabo, S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **2010**, *38*, e87. [CrossRef] [PubMed]
39. Llamas, B.; Holland, M.L.; Chen, K.; Cropley, J.E.; Cooper, A.; Suter, C.M. High-resolution analysis of cytosine methylation in ancient DNA. *PLoS ONE* **2012**, *7*, e30226. [CrossRef]
40. Pedersen, J.S.; Valen, E.; Velazquez, A.M.; Parker, B.J.; Rasmussen, M.; Lindgreen, S.; Lilje, B.; Tobin, D.J.; Kelly, T.K.; Vang, S.; et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **2014**, *24*, 454–466. [CrossRef]

biology

MDPI

Article

Exploring the Functionality of Mesh-to-Mesh Value Comparison in Pair-Matching and Its Application to Fragmentary Remains

Zoe McWhirter [1], Mara A. Karell [1,2], Ali Er [3], Mustafa Bozdag [3], Oguzhan Ekizoglu [4] and Elena F. Kranioti [5,*]

1 Edinburgh Unit for Forensic Anthropology, University of Edinburgh, Edinburgh EH8 9AG, UK; zmcwhirter9@gmail.com (Z.M.); ad7900@coventry.ac.uk (M.A.K.)
2 School of Psychological, Social and Behavioural Sciences, Faculty of Health and Life Science, Coventry University, Coventry CV1 5FB, UK
3 Department of Radiology, Health Sciences University, Tepecik Training and Research Hospital, Izmir 35180, Turkey; alier1717@yahoo.com (A.E.); bozdagmustafa.84@gmail.com (M.B.)
4 Center of Legal Medicine, University of Laussane-Geneva, 1015 Lausanne, Switzerland; drekizoglu@gmail.com
5 Forensic Medicine Unit, Faculty of Medicine, University of Crete, 700 10 Heraklion, Greece
* Correspondence: ekranioti@uoc.gr; Tel.: +30-(2810)-542097

Simple Summary: Forensic anthropologists often face the task of analysing a mixed group of skeletal remains or matching a solitary bone with the rest of a skeleton to determine if it belongs to the same individual. One of the best ways to do this is by pair-matching left and right bones of the same type. Common pair-matching methods experience issues such as high levels of subjectivity, lack of reliability, or expensive cost of implementation. This study explores the application of the relatively new method, mesh-to-mesh value comparison (MVC), which matches paired bones based on morphological shape to determine the likelihood that they derive from the same individual. This study sought to expand on the success found in past publications using MVC and to see how well it performed on a sample of clavicles, a bone known for having a high degree of bilateral variability, of 80 modern Turkish individuals. This study also explored whether MVC can reliably match fragmented bones to their intact counterpart. Results show MVC successfully matched 88.8% of paired clavicles and suggest the method continues to be a promising avenue for pair-matching that is not affected by ancestry and may be applicable to fragmented remains with further study.

Abstract: Many cases encountered by forensic anthropologists involve commingled remains or isolated elements. Common methods for analysing these contexts are characterised by limitations such as high degrees of subjectivity, high cost of application, or low proven accuracy. This study sought to test mesh-to-mesh value comparison (MCV), a relatively new method for pair-matching skeletal elements, to validate the claims that the technique is unaffected by age, sex and pathology. The sample consisted of 160 three-dimensional clavicle models created from computed tomography (CT) scans of a contemporary Turkish population. Additionally, this research explored the application of MVC to match fragmented elements to their intact counterparts by creating a sample of 480 simulated fragments, consisting of three different types based on the region of the bone they originate from. For comparing whole clavicles, this resulted in a sensitivity value of 87.6% and specificity of 90.9% using ROC analysis comparing clavicles. For the fragment comparisons, each type was compared to the entire clavicles of the opposite side. The results included a range of sensitivity values from 81.3% to 87.6%. Overall results are promising and the MVC technique seems to be a useful technique for matching paired elements that can be accurately applied to a Modern Turkish sample.

Keywords: forensic anthropology; MVC; 3D modelling; pair-matching; computed tomography; fragmentation; clavicle

Citation: McWhirter, Z.; Karell, M.A.; Er, A.; Bozdag, M.; Ekizoglu, O.; Kranioti, E.F. Exploring the Functionality of Mesh-to-Mesh Value Comparison in Pair-Matching and Its Application to Fragmentary Remains. *Biology* **2021**, *10*, 1303. https://doi.org/10.3390/biology10121303

Academic Editor: Andrés Moya

Received: 26 October 2021
Accepted: 6 December 2021
Published: 9 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Commingled assemblages and isolated skeletal elements are often encountered in the archaeological record as well as in contemporary forensic-related fieldwork [1,2]. The concept of commingled remains refers to a single context in which there is a mixing of fragmented or whole skeletal elements belonging to two or more individuals [3,4]. The definition of commingled assemblages can be further specified as a mixing of the remains to the degree in which further scientific study is necessary to differentiate the various components [4]. The commingled nature of the context can arise through a multitude of processes including animal scavenging, abiotic taphonomic processes, and human activity [4–6]. These atypical contexts provide unique challenges in determining the ideal method to sort and analyse the associated osteological material in the pursuit of answering important questions relevant to the study of past populations or forensic investigations [6,7]. Multiple individual burials have often been observed as a regular practice in the Paleolithic period and throughout history [2,4,8,9]. In the more recent past, mass killings have led to many forensic anthropologists encountering an increasing number of sites with commingled remains [4,6]. One of the primary steps for approaching these challenges is to quantify the skeletal elements, define the minimum number of individuals, and then re-associate as many of the skeletal elements as possible in order to individualise the sample [6,8,10].

The most commonly applied method for re-association is the visual examination for similarities in size, shape, and taphonomic changes in order to pair-match skeletal elements [11]. Despite its popularity and longstanding application, there are several limitations to the visual assessment method, most of which stem from the subjective nature inherent in its application. There is no way to standardise observations made by distinct observers and conclusions can be difficult to justify, something that would be a huge detriment to forensic contexts. The accuracy of results is also quite heavily varied as it depends almost entirely on the level of experience held by the individual carrying out the assessment [12].

Another approach often employed for the re-association of human remains is osteometric sorting, which is also concerned with the attempt to pair match left and right skeletal elements. Osteometric sorting can be defined as the "formal use of size and shape to sort bones from one another" [12] (p. 1) and relies upon the metric analysis of different bones and the application of statistical regression formulae to match them with other bones from the same individual [11,13]. The underlying concept is that the degree of robusticity and overall size will be similar amongst all skeletal elements belonging to the same individual. The technique makes an attempt to move beyond the subjective nature of visual assessment by employing statistical models and formulae in order to increase replicability amongst different observers as well as to provide an avenue for quantifying the differences between size and robusticity which would allow for stronger justifications to be made when publishing or presenting resulting pair match conclusions [11,12]. There are many benefits to the technique and include the low cost of utilisation, quick return of results, and low error rates [12]. While it is an improvement upon the previously discussed visual observation method and its heavily subjective nature, there are still many limitations that can be encountered in the use of osteometric sorting. One major limitation is the failure of the method to consider the bilateral asymmetry that may exist within an individual [12,14]. It is well-known that handedness and other factors affect the size and morphology of bones and thus it is erroneous to ignore the effects this asymmetry may have on the expression of robusticity and size within an individual [14]. Another situation in which osteometric sorting may fall short is when attempting to sort individuals of a similar size [12]. This can be a major limitation in a diverse range of settings including, but certainly not limited to, martial-related commingled contexts where most individuals are young adult males from a similar population [9].

While DNA testing is a proven method for re-associating elements, it is also extremely costly and time-consuming and many protocols for dealing with complex commingling include the sorting of remains utilising other less expensive methods prior to the eventual

application of DNA analysis, arguably making DNA a last, supplementary step to consider when sorting human remains instead of a primary, stand-alone method [15]. The level of preservation and the degree of taphonomic alterations are additional limitations in the use of DNA analysis for re-associating skeletal elements.

New methodologies employed virtual tools of re-association and pair matching. For example, in a relatively new study, the researchers utilised a sample of 111 metacarpals originating from 17 individuals to perform a pair-matching test. Two-dimensional photographs were utilised to place landmarks on the metacarpals. The hypothesis of the study was that "shape differences would be smaller in bones belonging to the same individual than in those belonging to different individuals" [16] (p. 120). The underlying concepts and theories behind the method are laudable and the consideration of ways in which shape can be quantified is extremely promising and intriguing for the future of pair-matching. Preliminary results showed a range of accurately identified pairs from 75.6% by one observer to 82.9% by the second observer with incorrect pairs made by both [16]. The major limitations of this technique would involve the small simple size, the overall lack of validation studies, the high degree of variability between observer rates of accuracy, and the slightly difficult to reproduce methodology.

Other novel methodologies have focused on the realm of 3D digital analysis in an attempt to overcome the shortcomings of the traditional 2D osteometric sorting method upon which they are based, specifically when applied to high degrees of bilateral asymmetry [17,18]. The first of which utilises digital 3D analysis techniques to compare the 46 variables including dihedral angles, cross-sectional area, and cross-sectional perimeter comparisons. The results showed true positive rates between 0.976 and 1.0 [17]. Similarly, Fancourt et al.'s [18] next-generation osteometric sorting uses 3D computer-automated analysis of data points forming a loop around the perimeter of a bone [18]. The authors found that the 3D analysis outperformed the original 2D osteometric sorting [18]. The promising result from both publications demonstrates the effectiveness of using 3D computerised methods to overcome shortcomings of pre-existing sorting methods.

Recently, a novel virtual method of pair-matching elements in commingled situations was proposed [19]. The mesh-to-mesh value comparison (MVC) method is based on the digital comparison of three-dimensional mesh geometries created from white light-scanned or computed tomography data of skeletal elements. This method has been employed with great success for pair-matching geometries of intact skeletal antimeres, that is, left and right sides, in humeri [19], parietal bones [20], and phalanges [21]. MVC is carried out by comparing the three-dimensional geometry of two skeletal elements and determining a numerical value which demonstrates the amount of similarity of the two elements [19]. The fundamental concept is that two paired elements belonging to the same individual will exhibit greater degrees of similarity than two elements belonging to different individuals. While this concept is not new and is a principal consideration in other pair matching techniques such as osteometric sorting and visual assessment, the traits MVC utilises to determine the similarity between bones is unique. The way the similarity values are generated in MVC is essentially by overlapping two bone models in the same three-dimensional space to determine the places in which the shapes differ and by how much. One of the novel features of MVC is that the method utilises all of the spatial data available and it does so in a three-dimensional landscape. This differs from the other pair-matching methods previously mentioned which focus on characteristics such as size or visual observations as well as from other geometric morphometric methods which rely on a limited number of specific landmarks on the bone as opposed to taking into consideration the entire external morphology and topography of the element in question, as MVC does [16]. MVC uses a "mesh-to-mesh" value which quantifies the difference between two meshes, or models, in millimetres; the lower a mesh-to-mesh value, the more similar the models are. The algorithms utilised to determine a mesh-to-mesh value are based on Iterative Closest Point (ICP) comparison algorithms [19].

Parietal bone pair-matching seemed to be the most successful with 98% sensitivity and 100% specificity [20], followed by the humeri with 100% sensitivity and 60% specificity [19]. Drawbacks on the method include the need for special skills in manipulating 3D data, building 3D models from scans, and securing mesh quality which makes the method time-consuming. Yet, the use of ROC analysis allows the method to be adjusted on the question at hand, that is whether two bones are more likely to belong to the same individual or if excluding that they do is the most probable outcome. This can be achieved by adjusting sensitivity and specificity levels.

Pre-existing methods of re-association commingled remains are varied and diverse. However, many are hindered by limitations such as a high cost of implementation, high degree of subjectivity, low level of accuracy, or a lack of validation studies confirming a proven, replicable accuracy rate of success [4,12,17,18,22,23]. Another important issue is that in many places, there are no available skeletal collections which can be utilised to develop or test these methods for a variety of reasons including ethical concerns, inability to macerate, excavate, or examine remains or due to the lack of documented material. Recently, studies utilising computed tomography (CT) scan data have become more popular and are viewed as a potential solution when physical skeletal material is inaccessible [24,25]. Specifically, there is a current need for techniques which can be accurately applied to the population of Turkey; The Human Rights Association in Turkey produced a report in 2014 discussing the location of 348 mass graves in Turkey containing the remains of 4201 individuals requiring analysis and identification [24] (p. 90). It is especially important that techniques employed by researchers involved in human rights-related excavations worldwide and regardless of time period are as accurate and cross-validated as possible due to the sensitive nature of the investigations. The use of CT scans from the contemporary Turkish population is an ideal approach to solve the current problem concerning the lack of anthropometric data in the country [24].

In this vein, the present study adopted the MVC methodology [19–21,26] to investigate its utility in pair-matching clavicles, a paired element that has received limited attention in pair-matching studies. In addition to developing a method for complete clavicles, the study aims to pair-match fragments for the first time, as these can be often encountered in commingled situations. The sample derives from Turkey and the development of a virtual method of pair-matching is an adequate fit for the application in mass graves in the lack of skeletal reference collections in the country as described above.

2. Materials and Methods

2.1. Sample

For this project, a total of 160 clavicles from randomly selected computed topography (CT) scans taken of 80 individuals were used (Table 1). The CT scan data utilised originates from Tepecik Training and Research hospital in Turkey, were taken in 2016 for a different project, and the files were anonymised prior to receipt by the researcher. The CT scans are in radiological position and were performed using a 64 slice CT scanner (Siemens Solutions, Erlangen, Germany). The scanning parameters are 80 kV, 115 mAs, with a slice thickness of 1 mm and 512 × 512 matrix.

Table 1. Biological information of the Modern Turkish sample used in this study.

Sex	Number (Total)	Healed Fractures	Under 28 Years
Male	54	4	6
Female	106	5	10
Total	160	9	16

The entire sample included 27 males and 53 females. Ages ranged from 15 to 65 with an average age of 42.5 years. There were eight individuals under the age of 28. The sternal epiphysis of the clavicle does not completely fuse until age 23 for females and 25 for males while visibility of the epiphyseal scar may remain until age 27–29 [27]. Nine of

the clavicles in the sample displayed evidence of healed fractures. These were deliberately included in the sample for comparison and results were analysed with the 9 fractured clavicles included as well as with them removed to determine the effect it would have on the attempted pair-matching.

2.2. Methods

2.2.1. Segmentation

3D models were created with semi-automated segmentation using the Amira 5.2.2 software package following a modified version of that described by Karell et al. [19] in the first publication of the MVC method. Figure 1 illustrates a model of a left clavicle.

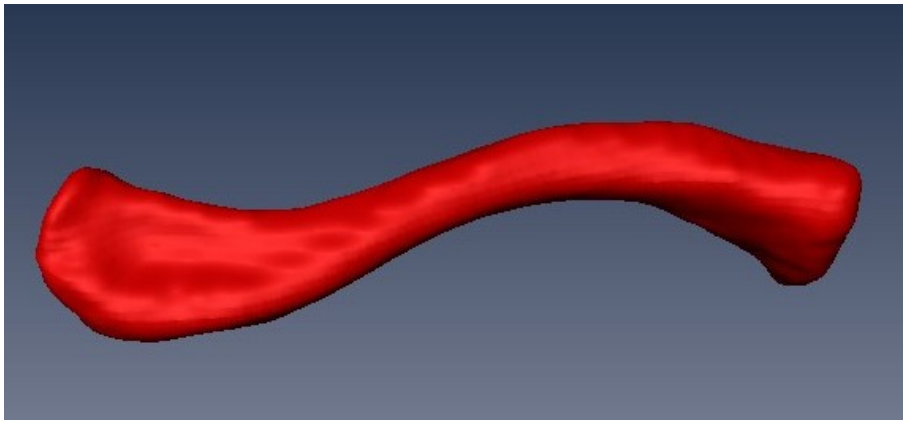


Figure 1. A completed model of the left clavicle belonging to individual 82. Created with Amira 5.2.2.

2.2.2. Simulation of Fragments

Following the segmentation process in Amira, the interior of the model was filled using the Fill Holes tool found in the Segmentation Editor. Once this was completed, the models were randomly cropped within the segmentation editor to create three different types of fragments; a fragment of the region adjacent to and including the medial epiphysis, which will be referred to as the sternal fragments, one consisting of portions of the midshaft, referred to as midshaft fragments, and one including the lateral epiphysis which will be referred to as the acromial fragments. This action was carried out for all 160 clavicles to create 160 models of each fragment type (480 in total) as seen in Figure 2.

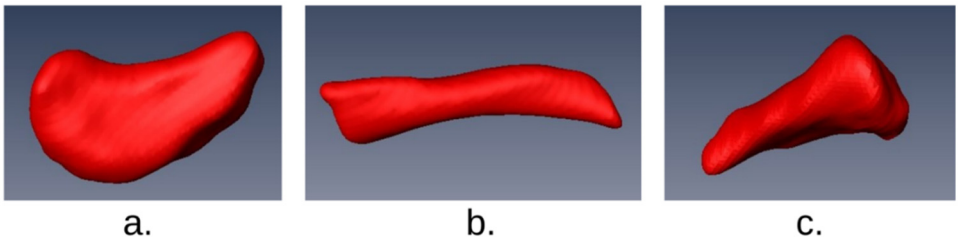


Figure 2. Examples of simulated fragments created using Amira 5.2.2.: (**a**) acromial fragment type, (**b**) midshaft fragment type, (**c**) sternal fragment type.

2.2.3. Mirroring

Following the segmentation and creation of the three-dimensional models, the right clavicles were imported into the Autodesk Netfabb software package and mirrored to create mirrored-rights. This step was carried out to ensure that all models can be appropriately compared once imported into the Viewbox 4.1 beta software.

2.2.4. Alignment

Once all of the right sided models were mirrored, all clavicle models were aligned using the Flexscan 3D software. First, the models were manually aligned as closely as possible. Once they appeared to all be in the same three-dimensional space and orientated in the same direction the alignment and fine-alignment actions were applied to the set. Following this step, the models were exported individually as OBJ files. The purpose for this step in the overall process is to eliminate any three-dimensional distance between the models and serves to reduce the amount of time the alignment step takes during the Viewbox Mesh comparison analysis.

2.2.5. Hollowing

Following the Flexscan alignment process, the models were subjected to a "hollowing" procedure using the Viewbox 4.1 beta software. That is the removal of any internal information and keeping only the external surface data for analysis. The nature of the mesh-to-mesh comparison involves only the morphological shape of the exterior surface of the bone models which makes the internal data irrelevant. Hollowing the models serves the purpose of reducing the amount of data that will need to be processed in the mesh similarity comparison process which will help to reduce the overall computing time. The average amount of data removed from each model was 27%.

2.2.6. Mesh-to-Mesh Value Comparison Using Viewbox

Following the previously described methods for creating and preparing the models, the sample was analysed to generate a mesh comparison value using the Mesh Similarity Tool in the Viewbox 4.1 beta software package. The mesh-to-mesh value is defined as the square root of the mean distances between the points of the two meshes.

The foundational algorithm utilised in the mesh comparison process within Viewbox 4.1 beta is a Trimmed Iterative Closest Point (Trimmed ICP) algorithm. Trimmed ICP has been lauded as a particularly useful moderation of the original ICP which performs well when conditions of three-dimensional comparisons involve the presence of shape defects and measurement outliers [28].

To compare all of the left and right models a folder was created with all models together and selected as the 'Mesh Folder'. A random model was selected as the reference mesh and the option to 'compare all meshes in mesh folder to each other' was chosen. Once all the proper parameters were set the mesh similarity was calculated and completed with a processing time of 21 h and 14 min; however, this time did not require any active input by the user.

Comparisons were carried out for the left and right clavicles, as well as comparisons of the fragmentary models to the complete clavicle models of the opposite side. Once the mesh values were generated for each sample, the generated Excel spreadsheets with the comparison values were used to perform two types of analysis in order to determine sensitivity and specificity values.

2.3. Mesh Value Analysis

2.3.1. Lowest Common Value Comparison

The lowest common value comparison method utilises a matrix method for selection in which the lowest mesh-to-mesh values for both the left and right sides must agree in order for a match to be determined. This method was developed by the authors of the original publication about mesh-to-mesh value comparison as an alternative to the previously attempted method of determining a threshold value to use in order to determine matches. As discussed by Karell et al. [19], the use of the cut off threshold value plus two standard deviations did capture almost all of the true matched pairs; however, it also included 51 values that were not true matches. Thus, an improved method for analysis was determined to be necessary [19]. The alternative method was shown to be a better method for selecting true pairs, at least for the humeri in the study. The lowest value comparison

method utilises a matrix method for selection in which the lowest mesh-to-mesh values for both the left and right sides must agree in order for a match to be determined. The benefit of this is that there should, in theory, be fewer false matches made.

The process of carrying out lowest value comparison method is executed within Microsoft Excel. This process involves formatting the Viewbox 4.1 beta produced results spreadsheets to determine the lowest three matches for each comparison. Through the use of sorting, macros, and relative references, the lowest agreed upon match by both paired elements is determined and a determination is made whether each row and column match is a true positive, true negative, false positive, or false negative.

A true positive value indicates that the value has been selected as the true match by Viewbox 4.1 beta and is also a known true match based on known sample data. A true negative will be a row in which there are no values selected and there is also no known true match for the model. For the purposes of this study, true negatives were only possible once data were intentionally deleted as original CT scan data were 100% paired. Thus, 20% of the results of each comparison sample were randomly removed to create a portion of true negatives. A false positive is a value in which the comparison method has selected a cell as containing a match but based on previous sample knowledge it is not a true pair. A false negative is when a model is not matched to any other model through the lowest value comparison process but does in fact have a true match.

Following the determination of all rows and columns, all the determinations were used to calculate sensitivity and specificity. Sensitivity was calculated as follows:

$$Sensitivity = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

Specificity was calculated using the following equation:

$$Specificity = \frac{True\ Negatives}{(True\ Negatives + False\ Positives)}$$

2.3.2. Receiver Operating Characteristics (ROC)

A ROC curve is a plot in which the sensitivity is plotted in function of the 100% specificity rate at different cut-off points of a specific parameter [29–31]. The plot of a ROC curve allows for the area under ROC curve (AUC) to be calculated. The AUC is a value which measures the success rate a specific parameter has when differentiating between two groups. For the purposes of mesh-to-mesh value comparison, this means that the AUC indicates how well the MVC method would perform with pair-matching. Through the creation of a ROC curve graph, it is possible to determine sensitivity and specificity values. The relationship between sensitivity and specificity is important when it comes to the analysis of ROC curves. A ROC curve of a test which has a perfect discrimination with a sensitivity and specificity of 100% will pass through the upper left corner of the graph.

3. Results

A total of 640 models, 160 intact clavicles, and 480 simulated fragments were compared and assessed to determine sensitivity and specificity using both variations of statistical analysis of the MVC method. Results are presented in Table 2.

3.1. Lowest Common Value Comparison

3.1.1. Entire Clavicle Models

To determine how well the automated version of the MVC method carried out the pair matching comparison for the clavicle models, two different methods of analysis were performed. The first method of analysing results is known as the lowest value comparison method and was carried out using Microsoft Excel.

This method was developed by the authors of the original publication about mesh-to-mesh value comparison as an alternative to the previously attempted method of deter-

mining a threshold value to use in order to determine matches. As discussed by Karell et al. [19], the use of the cut-off threshold value plus two standard deviations did capture almost all of the true matched pairs. However, it also included 51 values that were not true matches, thus an improved method for analysis was determined to be necessary [19]. The alternative method was shown to be a better method for selecting true pairs, at least for the humeri in the study. The lowest value comparison method utilises a matrix method for selection in which the lowest mesh-to-mesh values for both the left and right sides must agree in order for a match to be determined. The benefit of this is that there should, in theory, be fewer false matches made.

The process of carrying out lowest value comparison method is executed within Microsoft Excel. This process involves formatting the Viewbox 4.1 beta-produced results spreadsheets to determine the lowest three matches for each comparison. Through the use of sorting, macros, and relative references, the lowest agreed upon match by both paired elements is determined and a determination is made whether each row and column match is a true positive, true negative, false positive, or false negative.

The analysis of the 160 complete clavicles utilising the lowest common value comparison method yielded a sensitivity of 88.8% and specificity of 42.5% (Table 2).

To determine the impact of pathology and age, separate analyses were performed. A sample of 144 models with the pathological specimens included but the under-28 individuals excluded was analysed and yielded a sensitivity of 81.8% with a specificity of 0% due to the absence of any true negatives. Similarly, a sample of 151 models was analysed with the under-28 clavicles included while excluding the pathological specimens, which yielded a sensitivity of 82.8% and a specificity of 26.1% (Table 2).

3.1.2. Simulated Fragment Models

The comparisons for the sternal fragment type yielded a sensitivity of 65.4% and a specificity of 52.6%. The acromial fragment type produced a sensitivity of 54% and a specificity of 40%. The midshaft fragment type comparison produced a sensitivity value of 31.3% and specificity of 37.8% (Table 2).

3.2. ROC Analysis

3.2.1. Entire Clavicle Models

A ROC curve analysis of the data containing the match mesh-to-mesh values for the total sample of 160 entire clavicles produced an AUC value of 0.94 with a standard error of 0.0131 and a *p*-value of <0.0001 (Figure 3a). The sensitivity was 87.6% and the specificity was 90.9% (Table 2).

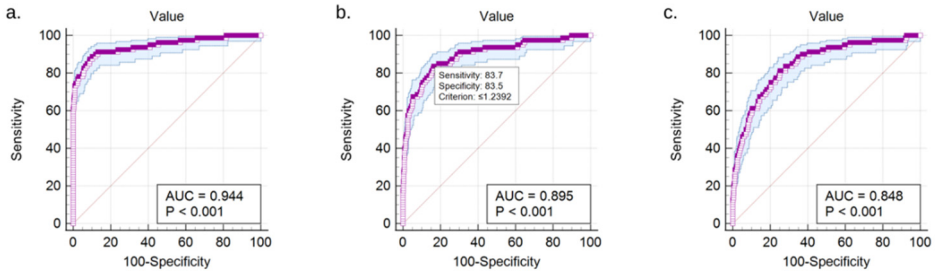


Figure 3. ROC curve diagram for (**a**) total sample of 160 clavicle models, (**b**) left sternal fragments matched to right entire clavicle models, and (**c**) left midshaft fragments matched to right entire clavicle models.

The ROC analysis of the data containing the entire clavicles with the healed fractures removed yielded an AUC value of 0.953 with a standard error of 0.0106 and a *p*-value of <0.001. The sensitivity was 89.5% and the specificity was 90.1%.

A separate ROC analysis performed on the sample of entire clavicles with the models belonging to individuals under the age of 28 removed produced an AUC of 0.940 with

a standard error of 0.0131 and a *p*-value of <0.0001. The sensitivity was 87.6% and the specificity was 90.98% (Table 2).

3.2.2. Simulated Fragment Models

ROC analysis performed on the results of the comparison of left sternal fragments to right clavicles produced an AUC value of 0.895 with a standard error of 0.0150 and a *p*-value of <0.001 (Figure 3b). The sensitivity was 83.8% and the specificity was 83.5%.

ROC analysis performed on the results of the comparison of left midshaft fragments to right clavicles produced an AUC value of 0.848 with a standard error of 0.0162 and a *p*-value of <0.001 (Figure 3c). The sensitivity was 81.3% and the specificity was 74.8%.

ROC analysis performed on the results of the comparison of left acromial fragments to right clavicles produced an AUC value of 0.934 with a standard error of 0.0132 and a *p*-value of <0.001. The sensitivity was 87.5% and the specificity of 87.9%.

Table 2. Results of all comparisons analysed in this study using both LCV and ROC statistical methods.

	LCV		ROC	
	Sensitivity	**Specificity**	**Sensitivity**	**Specificity**
160 clavicles	88.8%	42.5%	87.6%	90.9%
151 clavicles (Pathological specimens excluded)	82.8%	26.1%	89.5%	90.1%
144 clavicles (Under age 28 excluded)	81.8%	0%	87.6%	90.98%
160 acromial fragments	54%	40%	87.6%	87.9%
160 midshaft fragments	31.3%	37.8%	81.3%	74.8%
160 sternal fragments	65.4%	52.6%	83.8%	83.5%

Figure 4a illustrates an example of a true match after aligning and comparing the two models (left, mirrored-right) using a colour map. Blue indicates small differences in shape while red indicates large differences. Figure 4b illustrates a mesh-to-mesh comparison of a non-pair.

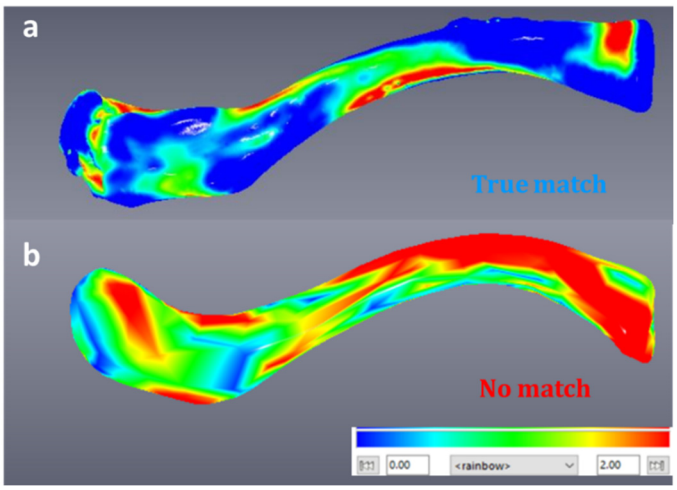


Figure 4. (**a**) True match of left and mirrored-right clavicle-visualisation of shape differences using a colour map in Viewbox beta software. (**b**) Visualisation of shape differences using a colour map for a left and mirrored-right clavicle that are not a true match.

4. Discussion

4.1. Comparisons with Other Studies and Methods

The human clavicle is one of the most variable bones in the skeleton in terms of morphological, anatomical, and biomechanical characteristics and has been described

as "non-conformist" [27,32,33]. Not only are the clavicles between different individuals extremely diverse but studies have noted a high degree of bilateral asymmetry amongst clavicles belonging to the same individual [27,34]. Clavicles have been extensively studied for several reasons, the most notable being the high rate at which it survives in a good degree of preservation due to the high proportion of compact bone as well as the utility of the medial epiphysis in terms of estimating age at death extending into the third decade of life [27,35,36]. For these reasons, the clavicle was selected to be the focus of this study.

One of the primary aims of this work was to determine the degree of success that can be expected when applying the automated version of mesh-to-mesh value comparison to pair-matching clavicles. The only studies published, to date, on this method (Table 3) are on humeri [19], temporal bones [20], mandibular fossae and condyles [26], and phalanges [21].

Table 3. LCV results of this study compared to previous MVC publications by Karell et al. (2016, 2017) and Acuff et al. (2021).

Sample	Author	Sensitivity	Specificity
45 mixed ancestry humeri (24 individuals)	Karell et al., 2016	95%	60%
120 Modern Greek temporals (60 individuals)	Karell et al., 2017	98%	100%
70 Cretan mandibular condyles (35 individuals)	Acuff et al., 2021	88.58%	0%
69 Cretan mandibular fossae (35 individuals)	Acuff et al., 2021	91.17%	100%
160 Modern Turkish clavicles (80 individuals)	This study	88.8%	42.5%
160 acromial fragments (80 individuals)	This study	54%	40%
160 midshaft fragments (80 individuals)	This study	31.3%	37.3%
160 sternal fragments (80 individuals)	This study	65.4%	52.6%

When compared to the LCV results of the Karell et al. humeri and temporal studies, the degree of accuracy found in this study is notably lower [19,20] (Table 3). The rate of sensitivity for the automated version of MVC when applied to the sample of 45 humeri is 95% while the resulting sensitivity in this study is 88.8% when analysed with the lowest value comparison (LCV) method. This discrepancy is not wholly unexpected as the clavicle is a much more irregular bone than the humerus and is known for expressing a marked degree of bilateral asymmetry [25,27]. The results are still positive and continue to place the automated mesh-to-mesh value comparison among the more accurate methods for pair-matching.

The 2021 study applying MVC to mandibular condyles and fossae experienced similar results to this study when using LCV analysis, yielding a sensitivity of 88.58% for condyles and 91.17% for fossae. These results are very close to those yielded in the comparisons of 160 clavicles in this study which may suggest that mandibular epiphyses and clavicles both perform similarly in MVC comparisons.

A previous study exploring pair-matching phalanges using the MVC method yielded the most promising and thorough ROC analysis results [21]. In that study, the best pair-matching bone was found to be the proximal phalanx of digit 3 and they found a sensitivity of 87.5% and specificity of 92.4%. This is similar to the 87.6% sensitivity and 90.9% specificity yielded by comparing the entire clavicles in this study.

4.2. Analysis Method: LCV vs. ROC

A third primary intention of this study was to explore the differences between the two types of analysis, LCV and ROC, and to determine which performs better when applied to MVC results.

The first type of analysis considered, lowest common value comparison (LCV), has many benefits. The underlying concept is that the lowest match of both the left and right sided models must agree or else it is not determined to be a match. This is especially useful in situations where it is important not to falsely match elements. Additionally, LCV comparison is performed using Microsoft Excel making it a very accessible process as there are no highly specialised software packages which require advanced training or high purchase costs to complete the analysis.

The benefits of the ROC curve analysis are likewise numerous. With the creation of a ROC curve, various types of insight into the data are gained. The additional option to create an interactive dot diagram is extremely useful in certain situations. With the interactive dot diagram, it is possible to choose whether sensitivity or specificity is more important and adjust the threshold in order to determine which values fall below a certain sensitivity-specificity percentage. One situation in which calculating a ROC threshold would be useful is when it is possible to carry out DNA analysis following the MVC process. For example, a mesh-to-mesh value comparison could be undertaken utilizing an interactive ROC dot diagram with 100% sensitivity selected which would mean that the overall number of potential matches would be reduced to those that performed well in the MVC process but with 100% sensitivity, no potential matches would be missed. It would then be simple to perform a DNA analysis on all bones that fall under the line determined by the diagram and then use the results from that analysis to determine the actual true match. This would reduce both the monetary expense as well as the waiting time inherent in the process of carrying out DNA analysis by reducing the original number of elements sent for analysis. This approach would greatly expedite the process as it takes significantly less time to perform an MVC match test than to analyse the DNA of every element in a given assemblage in the pursuit of individualisation.

In addition to the previous benefits, the ROC curve analysis automatically utilises bootstrapping which results in a greater sample size, making accuracy results more reliable [29].Last, the ROC analysis is much less time consuming for the researcher than completing an LCV analysis and is something that could even be put into practice in the field or in situations where spending hours on the computer is not ideal or possible.

The ideal method for analysing results produced using the automated MVC method cannot be determined without consideration of the type of sample, situation, and expected result of study. In this study, both LCV and ROC performed similar in regard to sensitivity for the entire clavicle models while ROC performed significantly better for the fragment comparisons.

4.3. The Effect of Age and Pathology

In addition to the inherent morphology of the clavicle and the effect this may have on the overall success rate of MVC, there are other factors that may have affected the accuracy results in this study. One of these factors is the inclusion of clavicles exhibiting evidence of healed fractures in the sample. In an attempt to determine what effect the presence of these nine healed fractured clavicles may have on the overall study, a separate sample excluding pathological bones was prepared and analysed. The LCV analysis yielded results in which the sample without the fractured clavicles was slightly less sensitive while the ROC analysis produced the opposite results. However, both methods produced similar sensitivity and it can be argued that the difference in accuracy is negligible and thus the presence of healed fractures in the sample was not a major hinderance or a factor that seems to have made a great impact on the overall performance of the MVC method. These results are interesting as they imply that the presence of observable pathology is not something that must be greatly considered when employing the automated version of MVC. Skeletal

pathology is a factor that is a common issue for several osteological analysis methods. While the analysis of the effect of pathology in this study was not a key aim and is in no way a conclusive statement on the performance of MVC for pathological samples, these results open an interesting new avenue of future research into the abilities of MVC.

The late-maturing nature of the clavicle is another factor worth consideration when attempting to compare the results of pair-matching clavicles to other studies relying on more typical long or other bones [19–21,26]. The sample used in this study included eight individuals under the age of 28 years old. The medial epiphyseal scar can remain visible late into the third decade [27,37]. It is also a commonly reported issue, especially amongst observers with little experience in working with x-rays and CT scans, to miss the medial epiphyseal flake or to be unable to observe the signs of the epiphyseal scar when creating a 3D model of a clavicle [35].

While the clavicle models belonging to these younger individuals were included in the overall sample of 160 clavicles, separate ROC and LCV analyses were completed on a sample with those models in question removed. The results show that when relying on the LCV method of analysis, the inclusion of the clavicles belonging to younger individuals had a positive effect on the results as the sensitivity was approximately 3% greater in the sample where they were included. The results of the ROC analysis were almost identical amongst both the samples indicating that the inclusion of the younger models has essentially no impact when using ROC statistics for analysis. These results could indicate several potential conclusions. It is possible that the errors made during the segmentation process when attempting to observe the flakes or epiphyseal scars were minimal, that the individuals discussed happen to have clavicles that are distinct, and thus pair-matching performs well in their case or, most plausibly, that the sample is too small to have a marked effect on the results. While it is interesting to consider that age is not a factor that negatively affects the MVC process, it should be taken into consideration that a thorough exploration of this concept would require a greater sample of younger individuals.

4.4. The Effect of Fragmentation

The exploration of how the automated version of MVC handles the pair matching of fragmentary or incomplete remains was another key aim of this study. Since the application of MVC to incomplete remains has not been thoroughly explored by other researchers to date, that aspect of this study was highly exploratory in nature. A recent study by Acuff et al. applied the MVC method to isolated portions of bone using the mandibular condyles and fossae as a sample [26]. The difference in this study is that the MVC comparisons were made between clavicle fragments and their intact clavicle counterpart as opposed to matching bone fragments to other fragments consisting of the same isolated portion of the entire bone.

The most highly performing fragment type were the fragments which consisted of the area near the lateral epiphysis which are referred to as the acromial fragments in this study. The ROC analysis produced an overall sensitivity of 87.5% and specificity of 87.9% (Table 4) which is only slightly lower than the results of the entire clavicle comparison. The LCV analysis yielded significantly lower results, showing an overall sensitivity of 65.7% and specificity of 40%. These results indicate that the MVC method may have the potential to match fragmentary remains and suggests that future explorations of matching fragments would benefit from focusing on the ROC analysis as it performed significantly better than the LCV. Considering that the fragments are being compared to entire clavicles as opposed to other fragment types, the success is expected to be lower as there is a large portion of the bone which is absent and thus cannot be compared. The potential logic underlying the improved performance of the acromial fragments when compared to the other two types of fragments can be related back to the morphology of the clavicle. Studies have often found this lateral retrocurved section to be one of the most variable regions of the clavicle, making it much more diverse in shape than the midshaft or medial epiphyseal (sternal) sections [25].

Table 4. Summary table of MVC fragment comparison results.

Fragment Type	LCV Sensitivity	LCV Specificity	ROC Sensitivity	ROC Specificity
Sternal	55.4%	56.5%	83.8%	83.5%
Midshaft	40.5%	37.8%	81.3%	74.8%
Acromial	65.7%	40%	87.5%	87.9%
Entire clavicles	88.8%	42.5%	87.6%	91.1%

The second-best performing fragment type was the medial epiphysis area which, in this study, was referred to as the sternal fragment type. The sternal fragments produced an average LCV sensitivity of 55.4%, specificity of 56.5%, and a ROC sensitivity of 83.8% and specificity 83.5% (Table 4). These results are still positive as even the lower performing LCV method yielded a percentage greater than 50% while the ROC results are more promising.

By far the worst performing fragment type was the ones that are made up on the central aspect of the clavicle and referred to as the midshaft fragments. The resulting ROC analysis yielded a sensitivity of 81.3% and specificity of 74.8% while the LCV yielded a sensitivity of 40.5% and specificity of 37.8%. While the ROC results are still positive, the LCV comparison results are extremely low and less significant than random chance when it comes to determining a true match. The possible logic behind the poor performance of the midshaft fragments is the fact that there is very little variation in shape amongst this area of clavicles. Unlike either epiphyseal area, there are also few notable bony landmarks which aid in the creation of a diverse or unique shape.

5. Conclusions

Pair-matching skeletal elements with the goal of re-associating remains to individualise skeletons is one of the most useful approaches to the study of commingled or isolated contexts involving human osteological material. While traditional methodology can be complex and varies greatly between situations, the pre-existing techniques have been proven to be lacking and are often difficult to reproduce between observers or highly dependent on the subjectivity of the researcher. Innovative new methods have investigated the incorporation of machine learning algorithms, computer software, three-dimensional modelling, and increasing utilisation of statistical formulae to combat the issues faced by pre-existing techniques [16–19,38]. The continual improvement in methodology available for the approach to sorting commingled assemblages is vital as multiple individual contexts are increasingly encountered by both osteoarchaeologists and forensic anthropologists. Thus, the expectation of the degree of accuracy and support for any conclusions made by research carried out in the field of osteological analysis as a whole continues to increase [6].

It is commonly acknowledged that the degree of accuracy involved in creating a biological profile of an individual skeleton tends to be greater when techniques that are either population-specific or shown to be unaffected by ancestral background are employed, making the population-specific validation of methods for analysis exceedingly critical. Through the course of this study, the MVC method for pair matching skeletal elements was analysed and attempts were made to validate its application to a contemporary Turkish sample of 160 three-dimensional clavicle models originating from computed tomography (CT) scans of 80 individuals of mixed age and sex. The overall results did not negate any of the claims made in the original publication and provide further evidence that the MVC process is a promising technique to employ when confronted with large- or small-scale commingled assemblages [19]. Fragmentary remains are often a roadblock when attempting to employ any method of analysis and this study hoped to determine whether that is indeed also the case for MVC. Results were mixed but promising and further research is necessary to determine the degree of accuracy that could be expected when attempting to pair match fragmentary or incomplete remains. This study also provided further support for the continued use of CT scan data as a stand-in for physical skeletal collections when

necessary and the positive effect this can have on the validation of methods for specific populations lacking in skeletal material available for research purposes.

Author Contributions: Conceptualisation, Z.M., M.A.K. and E.F.K.; methodology, Z.M., M.A.K. and E.F.K.; software, Z.M. and M.A.K.; validation, Z.M., formal analysis, Z.M.; investigation, Z.M. and M.A.K.; resources, O.E., A.E. and M.B.; data curation, O.E., A.E. and M.B.; writing—original draft preparation, Z.M.; writing—review and editing, M.A.K., O.E., A.E., M.B. and E.F.K.; visualisation, Z.M. and E.F.K.; supervision, E.F.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Tepecik Training and Research Hospital, Izmir (2016/32/11) and the Ethics Committee of the University of Edinburgh (2017).

Informed Consent Statement: Anonymised CT scans with demographic information was used for this study. Patient consent was waived due to the nature of the data.

Data Availability Statement: Data can be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adams, B.J.; Byrd, J.E. Resolution of small-scale commingling: A case report from the Vietnam War. *Forensic Sci. Int.* **2006**, *156*, 63–69. [CrossRef] [PubMed]
2. Osterholtz, A. *Theoretical Approaches to Analysis and Interpretation of Commingled Human Remains*, 1st ed.; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.
3. Lambacher, N.; Gerdau-Radonic, K.; Bonthorne, E.; de Tarazaga Montero, F.J.V. Evaluating three methods to estimate the number of individuals from a commingled context. *J. Archaeol. Sci. Rep.* **2016**, *10*, 674–683. [CrossRef]
4. GarridoVaras, C.; Intriago Leiva, M. *Managing Commingled Remains from Mass Graves: Considerations, Implications and Recommendations from a Human Rights Case in Chile*; Forensic Science International; Elsevier Ireland Ltd.: Dublin, UK, 2012; Volume 219, pp. e19–e24.
5. Wagner, S. *Commingled Remains, Commingled Human Remains: Methods in Recovery, Analysis, and Identification*; Elsevier: Amsterdam, The Netherlands, 2014.
6. Ubelaker, D. Approaches to the study of commingling in human skeletal biology. In *Advances in Forensic Taphonomy. Method, Theory and Archaeological Perspectives*; Haglund, W., Sorg, M., Eds.; CRC Press: Boca Ratón, FL, USA, 2002.
7. Byrd, J.E.; Legarde, C.B. *Osteometric Sorting, Commingled Human Remains: Methods in Recovery, Analysis, and Identification*; Elsevier: Amsterdam, The Netherlands, 2014.
8. Robb, J.; Elster, E.S.; Isetti, E.; Knüsel, C.J.; Tafuri, M.A.; Traverso, A. Cleaning the dead: Neolithic ritual processing of human bone at Scaloria Cave, Italy. *Antiquity* **2015**, *89*, 39–54. [CrossRef]
9. Holst, M.K.; Heinemeier, J.; Hertz, E.; Jensen, P.; Løvschal, M.; Mollerup, L.; Odgaard, B.V.; Olsen, J.; Søe, N.E.; Kristiansen, S. Direct evidence of a large Northern European Roman period martial event and postbattle corpse manipulation. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 5920–5925. [CrossRef]
10. Konigsberg, L.W.; Adams, B.J. *Estimating the Number of Individuals Remains: A Critical Evaluation of Methods, Commingled Human Remains: Methods in Recovery, Analysis, and Identification*; Elsevier: Amsterdam, The Netherlands, 2014.
11. Thomas, R.M.; Ubelaker, D.H.; Byrd, J.E. Tables for the Metric Evaluation of Pair-Matching of Human Skeletal Elements. *J. Forensic Sci.* **2013**, *58*, 952–956. [CrossRef] [PubMed]
12. Byrd, J.E.; Adams, B.J. Osteometric sorting of commingled human remains. *J. Forensic Sci.* **2003**, *48*, 717–724. [CrossRef] [PubMed]
13. Lynch, J.J.; Byrd, J.; LeGarde, C.B. The Power of Exclusion using Automated Osteometric Sorting: Pair-Matching. *J. Forensic Sci.* **2018**, *63*, 371–380. [CrossRef]
14. Steele, J.; Mays, S. Handedness and directional asymmetry in the long bones of the human upper limb. *Int. J. Osteoarchaeol.* **1995**, *5*, 39–49. [CrossRef]
15. Hines, D.Z.; Vennemeyer, M.; Amory, S.; Huel, R.L.; Hanson, I.; Katzmarzyk, C.; Parsons, T.J. *For DNA Analysis in Commingled Cases, Commingled Human Remains: Methods in Recovery, Analysis, and Identification*; Elsevier: Amsterdam, The Netherlands, 2014.
16. Garrido-Varas, C.; Rathnasinghe, R.; Thompson, T.; Savriama, Y. A New Method to Pair-match Metacarpals Using Bilateral Asymmetry and Shape Analysis. *J. Forensic Sci.* **2014**, *60*, 118–123. [CrossRef]
17. Bertsatos, A.; Chovalopoulou, M. Advances in Osteometric Sorting: Utilizing Diaphyseal CSG Properties for Lower Limb Skeletal Pair-Matching. *J. Forensic Sci.* **2020**, *65*, 1400–1405. [CrossRef] [PubMed]
18. Fancourt, H.S.; Lynch, J.J.; Byrd, J.E.; Stephan, C.N. Next-generation osteometric sorting: Using 3D shape, elliptical Fourier analysis, and Hausdorff distance to optimize osteological pair-matching. *J. Forensic Sci.* **2021**, *66*, 821–836. [CrossRef] [PubMed]

19. Karell, M.A.; Langstaff, H.K.; Halazonetis, D.; Minghetti, C.; Frelat, M.; Kranioti, E. A novel method for pair-matching using three-dimensional digital models of bone: Mesh-to-mesh value comparison. *Int. J. Leg. Med.* **2016**, *130*, 1315–1322. [CrossRef]
20. Karell, M.A.; Lay, M.; Langstaff, H.K.; Kranioti, E.F. Pair-matching temporals using a digital mesh-to-mesh value comparison method. *La Revue de Médecine Légale* **2017**, *8*, 185. [CrossRef]
21. Tsiminikaki, K.; Karell, M.A.; Nathena, D.; Halazonetis, D.; Spanakis, K.; Kranioti, E.F. Three-Dimensional Geometry of Phalanges as a Proxy for Pair-Matching: Mesh Comparison Using an ICP Algorithm. In *Biomedical Visualisation. Advances in Experimental Medicine and Biology*; Rea, P.M., Ed.; Springer: Cham, Switzerland, 2019; Volume 1205. [CrossRef]
22. Herrmann, N.P.; Devlin, J.B.; Stanton, J.C. *Assessment of Commingled Human Remains Using a GIS-Based and Osteological Landmark Approach, Commingled Human Remains: Methods in Recovery, Analysis, and Identification*; Elsevier: Amsterdam, The Netherlands, 2014.
23. Vickers, S.; Lubinski, P.M.; Henebry DeLeon, L.; Bowen, J.T., Jr. Proposed Method for Predicting Pair Matching of Skeletal Elements Allows Too Many False Rejections. *J. Forensic Sci.* **2015**, *60*, 102–106. [CrossRef] [PubMed]
24. Ekizoglu, O.; Er, A.; Bozdağ, M.; Akcaoglu, M.; Can, I.O.; García-Donas, J.G.; Kranioti, E.F. Sex estimation of the tibia in modern Turkish: A computed tomography study. *Leg. Med.* **2016**, *23*, 89–94. [CrossRef] [PubMed]
25. Lambert, S.; Al-Hadithy, N.; Sewell, M.D.; Hertel, R.; Südkamp, N.; Noser, H.; Kamer, L. Computerized tomography based 3D modeling of the clavicle. *J. Orthop. Res.* **2016**, *34*, 1216–1223. [CrossRef] [PubMed]
26. Acuff, A.S.; Karell, M.A.; Spanakis, K.E.; Kranioti, E.F. Pair-Matching Digital 3D Models of Temporomandibular Fragments Using Mesh-To-Mesh Value Comparison and Implications for Commingled Human Remain Assemblages. In *Biomedical Visualisation. Advances in Experimental Medicine and Biology*; Rea, P.M., Ed.; Springer: Cham, Switzerland, 2021; Volume 1317. [CrossRef]
27. Black, S.; Scheuer, L. Age Changes in the Clavicle: From the Early Neonatal Period to Skeletal Maturity. *Int. J. Osteoarchaeol.* **1996**, *6*, 425–434. [CrossRef]
28. Chetverikov, D.; Svirko, D.; Stepanov, D.; Krsek, P. The Trimmed Iterative Closest Point algorithm. In *Object Recognition Supported by User Interaction for Service Robots*; IEEE: Manhattan, NY, USA, 2003; Volume 3, pp. 545–548.
29. Ruopp, M.D.; Perkins, N.J.; Whitcomb, B.W.; Schisterman, E.F. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Bio. J.* **2008**, *50*, 419–430. [CrossRef]
30. Zweig, M.H.; Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561–577. [CrossRef]
31. Hanley, J.A.; McNeil, B.J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **1983**, *148*, 839–843. [CrossRef]
32. Sehrawat, J.S.; Pathak, R. Variability in anatomical features of human clavicle: Its forensic anthropological and clinical significance. *Transl. Res. Anat.* **2016**, *3*, 5–14. [CrossRef]
33. Daruwalla, Z.J.; Courtis, P.; Fitzpatrick, C.; Fitzpatrick, D.; Mullett, H. Anatomic variation of the clavicle: A novel three-dimensional study. *Clin. Anat.* **2010**, *23*, 199–209. [CrossRef]
34. Auerbach, B.M.; Raxter, M.H. Patterns of clavicular bilateral asymmetry in relation to the humerus: Variation among humans. *J. Hum. Evol.* **2008**, *54*, 663–674. [CrossRef]
35. Ekizoglu, O.; Hocaoglu, E.; Inci, E.; Can, I.O.; Aksoy, S.; Sayin, I. Estimation of forensic age using substages of ossification of the medial clavicle in living individuals. *Int. J. Leg. Med.* **2015**, *129*, 1259–1264. [CrossRef] [PubMed]
36. Loh, J.L.; Wong, K.L.; Hwang, S.; Shen, L.; Murphy, D.P.; Daruwalla, Z.J. Orthopedic asymmetry and clavicle length. *Clin. Anat.* **2015**, *28*, 964. [CrossRef]
37. Mohsin, A.; Alam, Z.; Faruqi, N.A. Bilateral variations in the growth and development of human foetal clavicle. *Biomed. Res. (India)* **2013**, *24*, 235–241.
38. Mahfouz, M.R.; Mustafa, A.; Fatah, E.E.A.; Herrmann, N.; Langley, N.R. Computerized reconstruction of fragmentary skeletal remains. *Forensic Sci. Int.* **2017**, *275*, 212–223. [CrossRef] [PubMed]

biology

MDPI

Review

Forensic Facial Comparison: Current Status, Limitations, and Future Directions

Nicholas Bacci *, Joshua G. Davimes, Maryna Steyn and Nanette Briers

Human Variation and Identification Research Unit, School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa; joshua.davimes@wits.ac.za (J.G.D.); maryna.steyn@wits.ac.za (M.S.); nanette.briers@wits.ac.za (N.B.)
* Correspondence: nicholas.bacci@wits.ac.za; Tel.: +27-11-717-2204

Simple Summary: Facial identification is an emerging field in forensic anthropology, largely due to the rise in closed circuit television presence worldwide, yet there is little published research in it. Our research group has conducted a series of studies testing the validity and reliability of the facial identification practice of morphological analysis. In this paper, we summarize the results of our studies and other latest advances in facial identification practice. In addition, we present a review of relevant technical literature on the limiting factors imposed on facial identification by closed circuit television, while making recommendations for practice and the future of this research niche based on a combination of our results and the technical know-how available. Facial identification research is a multidisciplinary task, with involvement from the field of anatomy, forensic anthropology, photography, image science, and psychology, among others. The value of this brief review is the bridging of these multiple disciplines to discuss the relevant needs and requirements of facial identification in forensic practice and future research.

Abstract: Global escalation of crime has necessitated the use of digital imagery to aid the identification of perpetrators. Forensic facial comparison (FFC) is increasingly employed, often relying on poor-quality images. In the absence of standardized criteria, especially in terms of video recordings, verification of the methodology is needed. This paper addresses aspects of FFC, discussing relevant terminology, investigating the validity and reliability of the FISWG morphological feature list using a new South African database, and advising on standards for CCTV equipment. Suboptimal conditions, including poor resolution, unfavorable angle of incidence, color, and lighting, affected the accuracy of FFC. Morphological analysis of photographs, standard CCTV, and eye-level CCTV showed improved performance in a strict iteration analysis, but not when using analogue CCTV images. Therefore, both strict and lenient iterations should be conducted, but FFC must be abandoned when a strict iteration performs worse than a lenient one. This threshold ought to be applied to the specific CCTV equipment to determine its utility. Chance-corrected accuracy was the most representative measure of accuracy, as opposed to the commonly used hit rate. While the use of automated systems is increasing, trained human observer-based morphological analysis, using the FISWG feature list and an Analysis, Comparison, Evaluation, and Verification (ACE-V) approach, should be the primary method of facial comparison.

Keywords: human identification; facial identification; CCTV; photography; forensic facial comparison; morphological analysis; FISWG; face mapping; disguises

Citation: Bacci, N.; Davimes, J.G.; Steyn, M.; Briers, N. Forensic Facial Comparison: Current Status, Limitations, and Future Directions. *Biology* **2021**, *10*, 1269. https://doi.org/10.3390/biology10121269

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 3 November 2021
Accepted: 1 December 2021
Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Cameras and photographic imagery have been used in surveillance, identification, and detection of criminals as early as the 19th century [1]. Anthropological standards have been used to depict portraits of regular criminals for law enforcement registries, similar to today's mugshot system. These registries were intended as a means for witnesses and

victims to conduct a facial review of potential suspects. However, the lack of standardization in image capture processes made these registries ineffective. The advent of judicial photography, in the late 19th century, incorporated anthropometry and relied on standardized conditions of image capture, featuring the well-known anterior and lateral facial views with neutral expression and stance [1,2] routinely used to this day by many police departments throughout the world. The facial anthropometry application was abandoned in favor of the more accepted fingerprint identification system [3], yet the facial image capture standards it relied on endured in facial depiction practices throughout the 20th century [1].

Depicting faces [1], facial anthropometry [2], and facilitating crime scene investigations [4,5] have relied on the use of photography in a forensic context almost since its development [1]. Probably the most recognized use of photography in a forensic setting, and its derivative in the form of video recording, is surveillance. Closed-circuit television (CCTV) was the natural progression of improved use of video technology that allowed for consistent monitoring and review of potential criminal activities [6]. CCTV surveillance systems have since the 1990s become increasingly more common and relied upon throughout the world [7–10] and are in fact considered by many communities the norm in public areas [11,12].

Deployment of CCTV surveillance is considered to act as a deterrent for local crime in monitored areas [8,13,14], often shifting criminal incidents to nearby unmonitored areas instead of completely eliminating them [10]. However, perhaps its most valuable contribution is its frequent use in criminal investigations [8,15]. An analysis of CCTV data in the United Kingdom showed that when CCTV data are available, criminal activity is substantially more likely to be resolved [15]. When the data were not of use, it was primarily due to its lack of availability or some fixed parameter of the surveillance system being suboptimal, such as the incident not being covered by CCTV, the system being faulty, or the images being of insufficient quality [15]. The criteria of usefulness of CCTV recordings vary greatly based on the intended use.

Other than general surveillance and criminal activity monitoring, facial examination is often of interest for the data extracted from many CCTV surveillance systems. This has become more evident as the deployment of CCTV systems and increases in crime have led to an increase in demand for facial identification [16–18]. This rise in demand is a direct outcome of the increased availability of image data, from both CCTV data [7,16] and photographic and video evidence from other sources, such as mobile phones [19].

Forensic facial identification falls under the discipline of facial imaging, which involves the use of visual facial data to assist the identification process [20]. Through the analysis of photographic or video evidence, forensic facial identification is routinely utilized to associate persons of interest to criminal activity [17]. Craniofacial identification involves multiple disciplines, such as facial approximation, facial composites and sketches, age progression and regression, photographic superimposition, molecular photofitting, facial depiction, and facial comparison [20]. Some of these techniques, such as facial approximation and facial composites and sketches, have been researched in some depth [20]. However, forensic facial comparison (FFC) for identification remains largely untested, despite its increasing demand [17,21].

Understanding that forensic facial comparison is a niche of research that needs further development requires the use of clear terminology. A colloquial confusion in terminology between facial identification and recognition is prominent throughout many discussions. This misnomer has been discussed by Schüler and Obertová [22], who clarified that identification is reliant on perfect agreement, which is different from recognition, understood as the innate psychological process humans employ at a glance to recognize a face, usually based on familiarity. Therefore, to attempt facial identification from a forensic anthropological perspective, a strict process of facial comparison is employed. Due to the innate process of recognition in any forensic facial comparison process, the distinction needs to be made clear. Recognition is employed generally as part of the investigative process of facial com-

parison and is holistic, rapid, and methodologically inconsistent with a high predisposition to error [23,24]. Identification, however, requires further systematic analysis involving standardized, detailed, comprehensive, and meticulously recorded methodology [22]. As such, forensic facial comparison must involve the human-based detailed examination of facial images for identity confirmation [25–27].

Another prominent misconception in facial identification (ID) involves the misuse of the term "facial recognition" to specifically refer to automated or semi-automated facial recognition systems, with this being fully adopted by many in the field of automated facial recognition (e.g., [28,29]). To avoid this miscommunication, certain studies refer to automated facial recognition as facial recognition technology (FRT) or systems [30]; however, this practice is not universally applied.

The misnomer of FRT and facial ID is often closely associated to the misconception of FRT being considered the ideal approach to facial ID. FRT systems apply a variety of computer-based methods to attempt confirmation of facial identity [29,31] and have proven high levels of accuracy in constrained circumstances [28,29]. While great advances have been achieved in the field of FRT [28,32], it remains associated with high false positive rates [32,33], strong racial biases [34], and other ethical concerns around privacy and consent that require resolution prior to the employment of FRT in a legal context. Most concerns revolve around the reliance of FRT systems on biometric information [35] and highly standardized images [36–38], which are often not available in the realistic unstandardized organization of most surveillance installations. As a result, while there are strong commercial and government incentives to deploy FRT systems, in part due to their large market share (USD 3.72 billion) [39], they are still reliant on human-based validation in their operating loops [40]. The need for human validation is further enhanced by the lack of varied databases used to develop and test these FRT systems [41]. Hence, until further varied and realistic databases are used to test and develop these FRTs, human observer-based facial image comparison is considered the preferred approach to facial ID [25,42–44] and will likely persist as the validation method of choice despite the improvement and widespread deployment of FRT systems.

Understanding the limitations and permissible applications of FRTs is crucial to conducting research in both FFC and FRT. The misconceptions and assumptions around FRT and FFC may pose a risk of driving researchers and funders away from conducting research in facial identification. This is primarily because most funders and new researchers would consider facial identification, and particularly FFC, as redundant in an era where FRT has become the norm. Despite these misconceptions, human-based facial identification methods, which are currently employed routinely in the judicial system, rely on forensic facial comparison [17,42].

Facial examination, also referred to as forensic facial comparison (FFC), must be applied using the Analysis, Comparison, Evaluation, and Verification (ACE-V) approach [27], commonly used in other forensic practices, such as fingerprint identification [45]. The ACE-V methodological approach is meant to integrate principles of the scientific method in forensic comparisons in order to enhance their implementation and reliability [45].

In the past, approaches to FFC included photo-anthropometry, facial superimposition, and morphological analysis (MA) [20,27], with morphological analysis being the currently accepted method as advised by both the Facial Identification Scientific Working Group (FISWG) (https://fiswg.org/index.htm accessed on 30 October 2021) and the European Network of Forensic Science Institutes (ENFSI) (https://enfsi.eu/ accessed on 30 October 2021) [27,46]. Application of MA relies on the detailed examination of specific facial features to reach a conclusion with regard to the similarity or dissimilarity of two or more faces [27]. The facial features are assessed subjectively, evaluated, and compared between the faces [27]. The selection of individual facial features often depends on the feature list utilized. Feature lists generally include both overall face composition and structure, individual anatomical feature components (e.g., hairline shape, ear helix morphology, nasal alae protrusion, etc.), and distinguishing characteristics such as scars, blemishes, piercings,

and tattoos (e.g., [47]). The current standard feature list used for facial comparison relies on criteria developed by the FISWG for facial comparison by MA [47]. An example of how this analysis is conducted is shown in Figure 1, using sample facial images from the Wits Face Database [41].

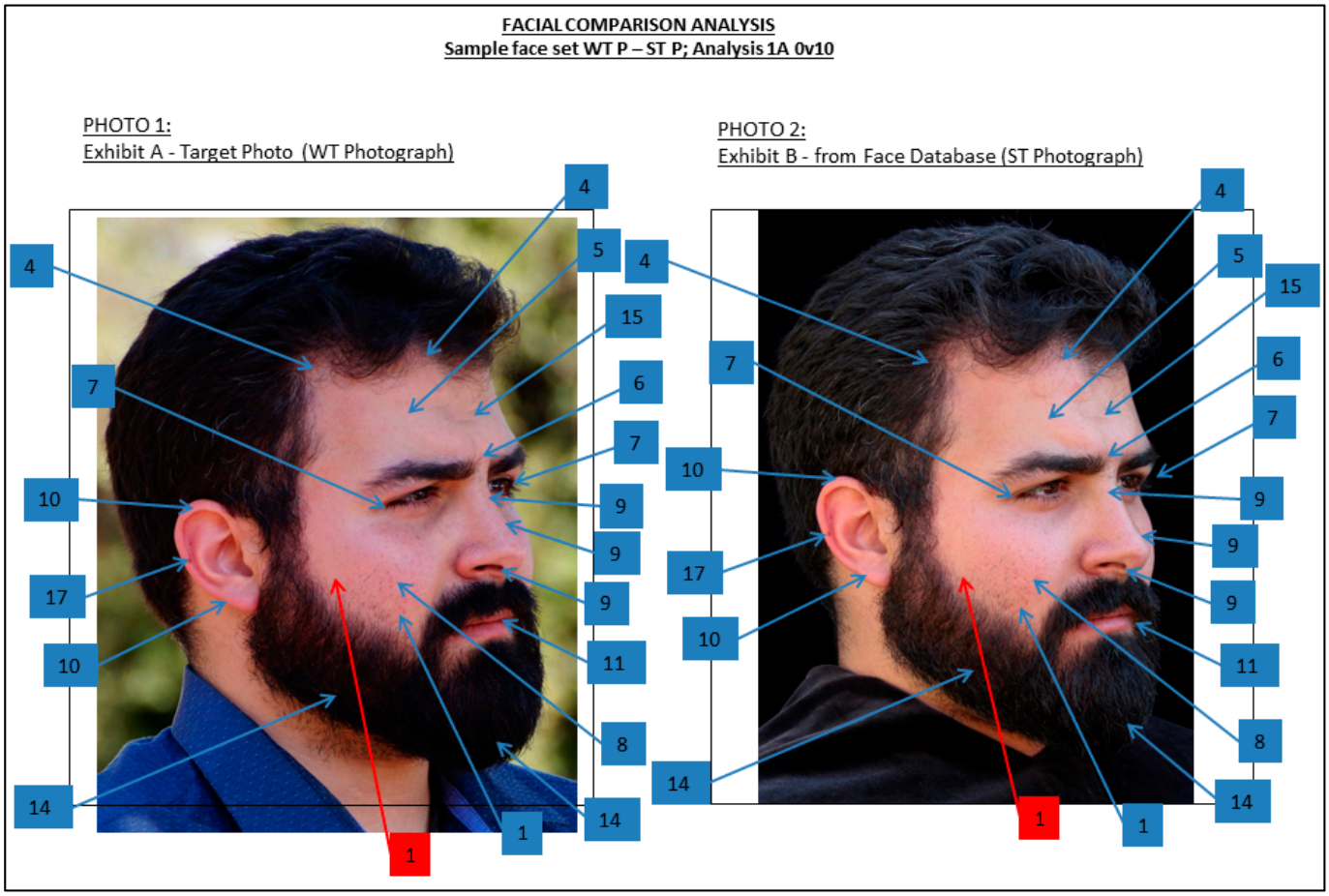


Figure 1. Example of a forensic facial comparison analysis process between a wildtype (WT) photograph and a standardized (ST) photograph from the Wits Face Database [41] sample images in the SAPS court chart format. The individual facial features are numbered, analyzed, compared, and evaluated between the two images using the FISWG feature list [47]. Features marked in blue indicate morphological similarity between the two images, while features marked in red indicate morphological dissimilarity. In the example provided, skin color appears different due to lighting discrepancies in the two images (red 1); however, skin texture appears similar (blue 1). The facial images used for Figure 1 are images of the corresponding author of the present manuscript and are part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The images can be found in the Wits Face Database data note, including the supplementary material for the Wits Face Database [41].

Recently, our research group (https://www.wits.ac.za/anatomicalsciences/hviru/ accessed on 30 October 2021) has conducted a series of validation studies to test the validity and reliability of FFC using the FISWG list (https://fiswg.org/index.htm accessed on 30 October 2021) of morphological features [21,41,48,49]. The aim of this paper is to summarize the results of these findings, thus elucidating the reliability and potential uses of FFC. Potential areas of caution and observed shortcomings are also discussed. Finally, recommendations as to the minimum standards for CCTV equipment are given, as well as guidelines for future directions in research.

2. Development of an African Facial Image Database

Although various facial databases exist (e.g., [50–61]), none of these were suitable for the systematic and blind testing envisaged for the purposes of the current stream of research on FFC validation. Some of these databases have small numbers of faces (e.g., [62,63]) or contain low-resolution images (e.g., [51,64]). As these databases were developed with different purposes in mind [65], and, with the exception of one [55], do not contain African faces, a new database was needed. A database containing African faces would also be invaluable in future research on the African continent.

Such a database was developed for the purposes of these studies, but due to the magnitude of such an undertaking, currently only males are included. This new Wits Face Database includes a total of 622 unique African male individuals aged between 18 and 35 at the time of recording, each with 10 photos associated to them, in five different views (anterior, left and right lateral, and left and right 45°) [41]. The 10 photographs were captured with high-resolution midrange cameras across two different conditions: a controlled setting with uniform background and obscured clothing at a subject-to-camera distance (SCD) of 1.5 m and an uncontrolled setting with a mixed background and visible clothing at an SCD of 5 m. This brought the total to 6220 facial photographs [41]. Out of the 622 participants, 337 (54.2%) were also recorded under different CCTV recording conditions [41]. The first group, recorded under a standard digital IP CCTV installation at approximately 3 m height, included 89 individuals; the second group, recorded at an eye-level digital IP installation (1.7 m installation height), included 76 participants; the third group included 107 participants, recorded by an older analogue CCTV installation (2.5 m height); and the last group, recorded by the same digital IP CCTV camera as the first group, included 34 and 31 participants wearing caps and sunglasses, respectively [41]. Throughout the CCTV data, large amounts of data loss were experienced, particularly with the internet protocol (IP) CCTV cameras, due to corruption, compression, and intermittent connectivity (Table 1).

Table 1. Composition of the Wits Face Database [41] CCTV data and detailed data loss experienced during database development as a result of the CCTV systems' technical limitations.

Database Cohort Organization	Unique Individuals	Photographs	Corresponding CCTV [1] Recordings	Data Loss (%)
ST [2] CCTV [1]—ST [2] Photographs	98	980	89	9.2%
Eye-level CCTV [1]—ST [2] Photographs	108	1080	76	29.6%
ST [2] CCTV [1] with Cap—ST [2] Photographs	45	450	34	24.4%
ST [2] CCTV [1] with Cap—ST [2] Photographs	41	410	31	24.4%
Total IP [3] CCTV [1] Data	292	2920	230	21.2%
Analogue CCTV [1]—ST [2] Photographs	111	1110	107	3.6%
CCTV [1] Grand Totals	403	4030	337	16.4%

[1] CCTV = closed-circuit television; [2] ST = standard; [3] IP = internet protocol.

While the inclusion of males only is a good step towards expanding the diversity in populations included in face databases, the non-existence of a female database remains a notable limitation to be aware of. In principle, the FISWG feature list should be generic enough to make it applicable across sex and population groups, but facial variations may potentially lead to variations in accuracies and reliability based on the biases and abilities of the observers. The existence of a within-group face recognition advantage (previously called own- or cross-race bias) has been well described and may play a role in the reported accuracies of FFC [66–71]. It is, therefore, essential that future databases include faces that are representative of all major populations. The newly developed database is now the largest African database of CCTV recordings and matching high-resolution facial photographs. It is available for all bona fide research that meets the criteria as set out by the Human Research Ethics Committee (HREC) (Medical) of the University of the Witwatersrand [41,72].

3. Outcomes of Validation Studies

Various standards exist worldwide as to how to express the levels of confidence when it comes to possible matches. In Australia, for example, facial comparison experts are expected to present evidence strictly in descriptive terms, which can lead to suggestive language, based on the expert's prejudice and opinion [26]. In England and Wales, FFC experts report on comparisons based on the Bromby scale of support [73], where the scales of support force experts to conclude whether two compared faces are a match regardless of image conditions or quality [26]. The Bromby scale is also inherently arbitrary with no clear distinction between each step of the scale. To alleviate these uncertainties, experts from the South African Police Services (SAPS) make use of a five-point scale that reflects the ability of an expert to analyze a given set of images, as well as the confidence level of a specific conclusion [17]. For application and testing, this scale was slightly adjusted to allow statistical testing to reflect an order of severity of conclusion. Namely, a score of 1 was assigned to confident positive identifications, a score of 2 to inconclusive identifications that showed some level of morphological similarity on certain specific facial features, a score of 3 that represented an inconclusive identification with overall holistic similarity of two faces compared, a score of 4 as a negative identification, and a score of 5 indicating impossible to analyze due to insufficient visibility of landmarks [21]. A visual overview of these outcomes is shown in Figure 2.

Morphological analysis on data derived from the newly developed Wits Face Database [41,72] using the FISWG feature list [47] was found highly accurate and reliable when comparing optimal standardized photographs to wildtype (informal) unstandardized photographs [21]. In an analysis of 75 sets of faces (each containing nine no match comparisons and one positive match comparison or 10 no match comparisons—compared to a target image, total $n = 750$ comparisons), the chance corrected accuracy and reliability were found to be almost perfect in optimal photographs (99.1% and 92.1%, respectively) [21]. In the analysis of 100 face sets ($n = 1000$ comparisons) with standard digital CCTV recordings as the target image, a lower accuracy (82.6%) and reliability (74.3%) were noted [21] (Figure 2). The lower performance of MA in standard CCTV was ascribed to the variation of conditions of the different equipment and its installation. Specifically, images obtained from the standard CCTV system were of poorer quality than the high-resolution controlled and wildtype photographs, due to a number of reasons. Firstly, the image resolution of the standard digital CCTV camera was lower (4MP) than that of the photographic cameras (18MP) [21]. Secondly, the CCTV field of view was broader and less focused on the face, partly due to the SCD being approximately 3 m. As such, a larger area was captured at a lower resolution, effectively reducing the actual resolution of the recorded faces [21]. Thirdly, between the CCTV camera and the captured face, an angle of incidence of 27° was formed, which appeared to limit visibility of the face, potentially shifting relative proportions of facial features [21]. The change in perspective and the limitations it placed on the facial comparison process likely contributed to the lower accuracy and reliability seen in the standard CCTV conditions [21].

Image lighting was also markedly different between photographs and CCTV recordings, making facial characteristics reliant on color (i.e., skin tone, luminescence, and color) redundant, since they appeared different even between matching images [21]. Variations in lighting also contributed to over-exposure of certain features, effectively limiting their utility in facial comparison [21]. Beyond these discrepancies and concerns, the almost perfect accuracies and the low false positive rates identified (<1.6%) (Figure 2) are encouraging for the use of MA in a legal context from both optimal photographs and standard CCTV installations [21].

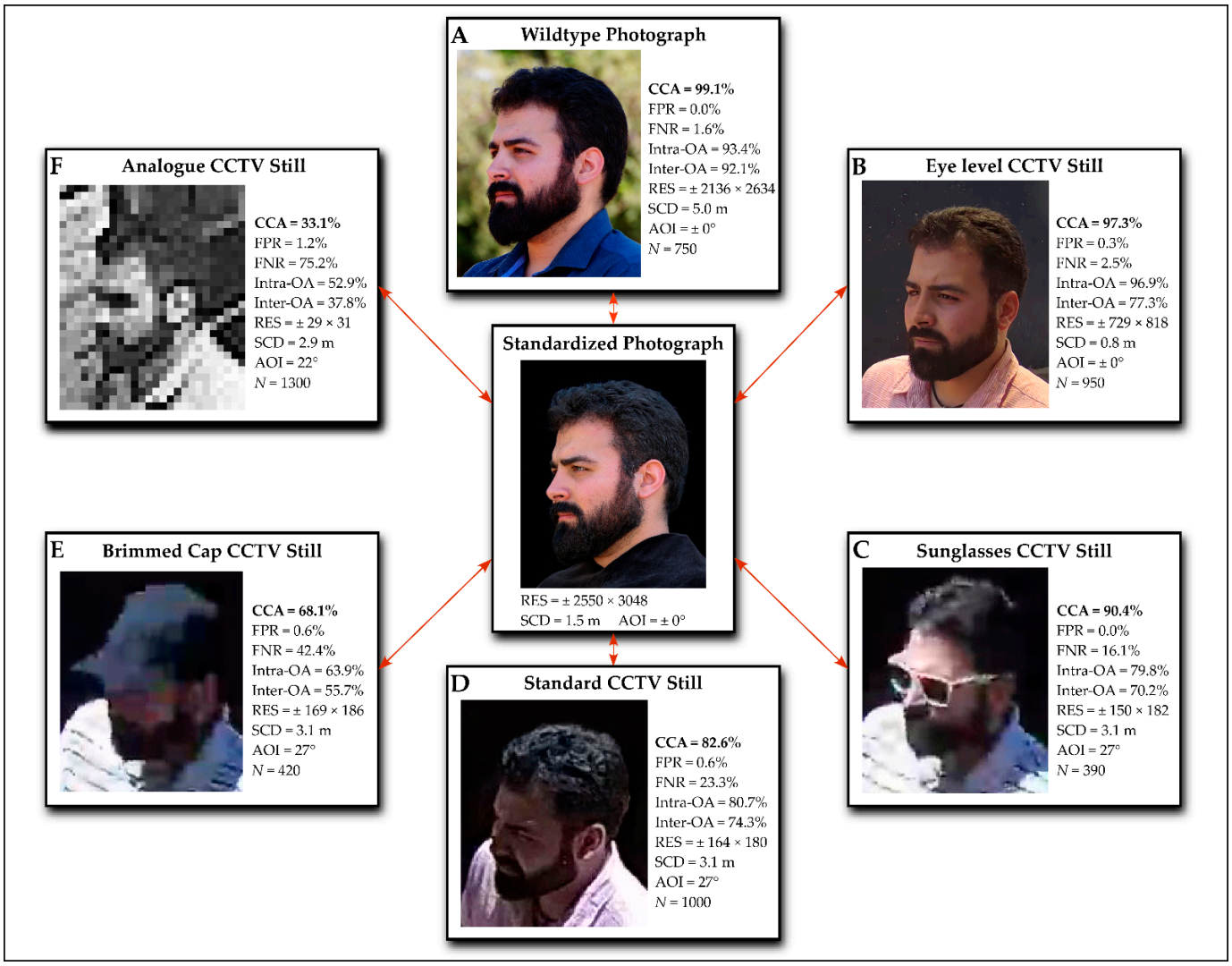


Figure 2. Visual summary of the validation studies testing morphological analysis across realistic photographic and CCTV conditions [21,48,49] using sample photographs and CCTV stills from the Wits Face Database [41]. Images (**A**) to (**F**) are samples of the target images from each set of conditions analyzed that were compared to the central image arising from the standardized photographs captured for each participant. All major statistical results and the details of the conditions of each comparison cohort are presented. Representative images of each condition are arranged from A to F in a clockwise order according to descending chance-corrected accuracy. The conditions of analysis were as follows: wildtype informal photographs (**A**) of similar quality to the standardized photographs; eye level digital CCTV still images (**B**); standard digital CCTV still images (**D**) with sunglasses (**C**) and with brimmed caps (**E**); and monochrome analogue CCTV still images (**F**). Key: CCA = chance corrected accuracy; FPR = false positive rate; FNR = false negative rate; OA = observer agreement; RES = resolution; SCD = subject-to-camera distance; AOI = angle of incidence; *N* = number of comparisons. The facial images used for Figure 2 are images of the corresponding author of the present manuscript and are part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The images can be found in the Wits Face Database data note, including the supplementary material for the Wits Face Database [41].

Following on the first set of analyses under fairly optimal conditions, a second set of tests was done on 130 face sets (n = 1300 comparisons), arranged as described above, recorded on a low-resolution suboptimal analogue CCTV system. The results were found to be much poorer, with accuracies as low as 33.1% with extremely high false negative rates (75.2%) and questionable reliability (37.8%) [48] (Figure 2). The contributing factors

to this decrease in accuracy were a pronounced angle of incidence (22°), lack of color, and particularly the low-resolution images [48]. However, determining which of these specific factors contributed the most to the low accuracy is not possible by study design, but the decreased quality of the images seems to be the most problematic factor [48]. The contribution of lacking color, however, is questionable, as facial examiners in certain countries conduct their comparisons in greyscale with the consideration that color can be considered misleading. This effect of color was also observed in a previous study, where attempting to match skin color between images proved futile due to lightning discrepancies between images [21]. Irrespective of the specific contribution, the combination of these factors was highly disruptive to the facial comparison analysis—even more so than the inclusion of disguises [49].

The above suboptimal comparisons were contrasted to 95 face sets (n = 950 comparisons) recorded at eye-level with a digital IP CCTV camera. As can be expected, eye-level digital CCTV images were found to yield better results than the standard CCTV installation [21,48]. An effective 0° angle of incidence and a much smaller SCD of 0.8 m seem to have simulated the most ideal CCTV conditions for facial comparison [48]. In fact, eye-level digital CCTV recording-based facial comparisons were almost as accurate (97.3%) and reliable (77.3%) as the standardized photograph to unstandardized photograph comparisons [21,48] (Figure 2). This outcome is telling of the factors that may have played the biggest role being angle of incidence and SCD, since the standard CCTV and the eye-level CCTV were identical cameras installed at different conditions [21,48]. However, to assess the extent of the influence these factors had on facial comparison, further targeted testing of these individual factors is required.

During the analysis of the data from the facial comparisons, two iterations were conducted—the strict and the lenient iterations. Under a strict iteration, only a confident positive identification was taken as a match, while under the lenient iteration, even inconclusive analyses with some morphological similarity in facial features were considered as matches along with the positive identification [17,21]. When reviewing the performance of MA in the analogue CCTV data, it was noted that a significantly altered performance resulted under different levels of analysis strictness. A strict iteration resulted in a worse performance in the analogue CCTV comparisons than across all other comparisons (photographs, standard CCTV, and eye-level CCTV) [48]. All other analyses from the various CCTV and photographic images showed improved performance under a strict iteration [21,48,49]. This outcome advocates that under particularly suboptimal conditions, such as analogue CCTV, even a strict approach to the analysis is ineffective in improving performance. However, the decreased accuracy under a strict iteration may be worth considering as a marker of suboptimal conditions. Effectively, when a strict iteration results in lower performance of MA in a particular dataset than a lenient iteration, that dataset should be viewed as being below a usable threshold for facial comparison. As such, recordings that perform worse in a strict iteration, particularly in cases where target exclusion is not possible, should be avoided for positive identification. Effectively, when testing the performance of MA in a given dataset extracted from a specific CCTV installation, both a strict and lenient iteration should be conducted. Should the strict iteration perform worse than the lenient iteration, then the specific CCTV installation that yielded that footage should be considered below a usable threshold for the purpose of FFC. This consideration of statistical analyses is included in our recommendations on how to conduct MA.

Across all of our studies, the best measure of accuracy was found to be the chance corrected accuracy (CCA) [21,48,49]. CCA was calculated by conducting a weighted Cohen's kappa (with squared weighting) on the assigned scores for each comparison contrasted to the actual true match-up information for each comparison trial. This is different to the normal hit rate or raw accuracy, which simply indicates the amount of correctly scored trials irrespective of the degree of error or the sample composition. This is also different to the balanced accuracy that is calculated when computing a confusion

matrix analysis, which is effectively the sum of the sensitivity and specificity divided by two [74]. In the studies' results, CCA varied the most and was seen as the most representative measure of accuracy, particularly when compared to the simple hit rate and balanced accuracy. These two accuracies presented skewed results towards true negatives due to the studies being conducted under a one-to-many comparisons context, an approach to facial analysis also seen as the harshest testing criteria for automated facial recognition systems [75,76]. As a result, the non-chance corrected accuracies appeared deceptively higher due to the high prevalence of true negative matches, despite other measures of performance indicating a more questionable outcome. With this consideration, future FFC studies should consider making use of CCA as their primary measure of accuracy as opposed to hit rate, historically the most common measure of accuracy.

Beyond the optimization of surveillance system installations specifically for facial comparison, an additional limiting factor investigated in these studies were the effects of disguises on facial comparison. We specifically investigated the effect of sunglasses (n = 390) and brimmed caps (n = 420) on FFC performance [49] (Figure 2). Overall, the performance of MA in faces disguised with sunglasses was markedly high (90.4%) [49], in fact surpassing the performance of facial comparison under the same standard CCTV conditions without sunglasses (82.6%), but not better than the photographic (99.1%) [21] or eye-level CCTV data (97.3%) [48] (Figure 2). This unusual consequence of sunglasses on facial comparison has also been observed by Davis and Valentine, who tested live subject to image identification [77]. These authors [77] suggested that the instruction that was given to participants conducting face matching tasks to rely on the external facial features with subjects disguised by sunglasses apparently increased their ability to recognize a face disguised by sunglasses. External facial features are in fact considered the most reliable set of features in unfamiliar face matching, as corroborated by other studies [78,79]. In FFC, conducted using the FISWG feature list, a methodical approach with a focus on all facial features including the external ones was followed. As a result, this methodical approach may have indirectly contributed to avoiding the limitation that sunglasses would normally pose on this comparison cohort. In contrast, faces disguised by brimmed caps yielded an exceedingly low CCA (68.1%) [49], yet not nearly as low as the analogue CCTV comparisons (33.1%) [48] (Figure 2). The limitations posed by brimmed caps appeared to have been compounded by the large angle of incidence of the standard CCTV recordings as well as the strong natural lighting from the sun. These two conditions, in conjunction with the brimmed caps, created shadows over the face, obscuring an even greater number of facial features, resulting in large-scale information loss [49]. This effectively rendered comparison much more difficult, as less than the lower half of the face and the ears could be evaluated [49].

Eyeglasses and various types of hats have historically been viewed as the most inconspicuous and common disguises [25,80,81]. Although the specific effects of various disguises have been discussed broadly, only one study has attempted applying MA to a disguised sample [49]. Despite their lack of testing in MA, in face matching recognition, brimmed caps were found to increase error rates over other comparison tasks [81]. Brimless caps and glasses, on the other hand, appear to have a less pronounced effect on match accuracy, varying by the method employed [82,83].

The success of MA in disguised faces was credited in large part to the FISWG feature list [47]. The use of even rudimentary feature instructions or even partial feature lists is able to increase the performance of facial comparison analyses [84–86], with a more pronounced effect noted for trained experts [84]. Our results from the disguised test of MA [49] reinforce these outcomes of other studies, further supporting the use of feature lists in MA.

4. Discussion

This paper summarized the outcomes of our recent studies testing MA and the FISWG feature list across varied conditions of facial CCTV images and photographs. In addition, it presented and discussed the major limitations of FFC. MA of faces, using a feature list, is accurate and valid, particularly when conditions are optimal (e.g., high-resolution photographs and high-resolution CCTV with limited perspective distortion/angle of incidence). Image quality had the most notable effect on facial comparison performance (analogue CCTV recordings), while brimmed caps were found to be the second-most limiting condition. Across both of these conditions, the major limiting factor appeared to be overall loss of facial feature information, with caps obscuring almost half of the face and the poor quality of analogue video material making most of the facial details indistinguishable.

4.1. Influence of CCTV Installations

To determine the minimum criteria for facial examination across various CCTV installations, a more thorough understanding of the conditions imposed on footage by specific installations is needed. This is of particular relevance with the continuing global increase in the installation and usage of CCTV systems that has been seen across private, public, and commercial sectors in the last two decades [7,46,87,88]. This increase can be attributed to multiple factors; however, two major drivers include advancements in computing and CCTV system production and a reduction in the associated costs [7]. The vast global increase in CCTV deployment has led directly to an increase in available data for use in potential criminal surveillance and related investigations.

While this global increase in CCTV data is beneficial to criminal investigation and facial comparison, there is a concerning lack of standardization of required installation, recording conditions, and image quality [20,82,89–92]. As a result, the usefulness of CCTV-derived facial images is difficult to assess and makes facial comparison challenging in contrast to controlled photographs and mugshots. These limitations along the CCTV imaging chain are often acknowledged; however, few studies have assessed their implication in facial comparison accuracies [21,48,82,92–94]. Successful facial identification assessment is hindered by inconsistent recording conditions and poor image quality. Facial comparison accuracy and data quality are, thus, directly correlated [95,96], especially in terms of individual accuracy variation across multiple analysts [97] and individual analyst ability overestimation [98].

CCTV camera placement is one of the major limitations in terms of recording conditions. Most surveillance systems are put in place in order to monitor large crowds or entry/access points and do not have FFC in mind. The placement of the camera is based on the field of view that can be monitored and is then complemented by the mounted height above ground. Camera height relative to subject distance gives the angle of incidence, and this is an important, and often detrimental, component for extracting facial details from recordings.

Typical surveillance camera mount heights are between 2.5 and 3 m on building exteriors and ceiling height for indoor surveillance [99]. The main justification behind these mounting heights is that it lowers the risk of cameras being vandalized, stolen, or obstructed. The problem with these standardized mounting heights is that they translate to a steep angle of incidence. This in turn reduces image quality and obscures relevant facial detail as a result of the increased SCD and subsequent loss of useable resolution [48,49]. This is particularly important in facial comparison, as the amount of visible facial features and the view in which the face is seen are crucial for successful identification [100]. People also tend to naturally tilt their heads inferiorly by 15–20° when walking [100], thereby further exacerbating this problem. The current recommended angle of incidence limit is 15°, as any steeper angle would result in significant loss of facial detail [100]. Eye-level mounted cameras at 1.8 m ground height provide an approximate 0–15° angle of incidence with the subject and provide the most optimal capture of facial detail even with natural head tilt [48]. Further individual variations in facial view, or pose, are additional factors

that may further affect facial comparison, even at eye-level CCTV placement, particularly under poor quality and capture conditions [48,101,102]. Our reports [48] show an overall better and more reliable performance of MA in a digitally captured sample (IP cameras) at eye-level height (1.7 m) compared with a suboptimal sample at an angle of incidence of 27° (mount height of 3.1 m) [48]. In contrast, our analogue CCTV data, captured at an angle of incidence of 22° and height of 2.5 m, performed worse overall, although it is unclear if this was because of camera position or poor image quality [48].

Camera placement, particularly in relation to positioned angle and mounted height, dictates the monitoring area, while the camera lens, its focal length, as well as the sensor, specifically its size and number of pixels, dictate the field of view, image quality, level of optical distortion, and noise present [99]. The distance between the camera and a subject or target will then affect the image composition, which directly affects target size on sensor or picture height, and level of perspective distortion [92,103,104]. All of these factors and components will affect the usefulness of an image for facial detection and subsequent comparison analysis. Monitoring a large crowd outside a building, for example, requires 5% of picture height, while detecting a specific target requires 10% [99]. A potential target must occupy more than 400% of screen height in order to conduct facial examination, and a minimum of 1 mm must be represented per pixel of the whole image (ISO62676 recommendations) [99]. Considering the conservative European standards for facial image comparison [105], a minimum of the top quarter of a subject must be included on screen height and the face would need to represent a minimum of 1000 pixels per meter of screen height [99]. As such, for each inch (2.5 cm) of a face represented in an image, a minimum of 25.4 pixels is required [99]. For this minimum pixel density to be maintained at set SCDs, certain lens focal lengths need to be utilized. For example, at a 5 m distance from camera to subject, a focal length of 4.2 mm is necessary on a $\frac{1}{2}$" sensor HD CCTV camera [99], which is considered a common IP camera type. The longer the focal length of the lens, the narrower the field of view; simultaneously, the smaller the camera sensor, the smaller the viewing angle and the higher the noise. Bigger sensors and higher pixel counts are, in theory, always better for security and forensic applications, especially in low-light performance; however, bigger optics are then also required, which increases camera size, weight, power, and most importantly, cost.

Lighting conditions can pose further challenges in recording optimal footage. Facial details may be lost to over- or under-exposure of a subject and may not be retrievable through post-processing [99]. In outdoor locations, the position of the sun and related shadows, the amount of ambient lighting based on time of day, or the combination of multiple light sources or reflective materials near the subject or camera all could lead to unbalanced exposure. This then ties to the sensitivity of the camera sensor and its dynamic range capability. Most modern IP cameras are better suited to handling high-contrast environments, but older analogue systems generally provide either over- or under-exposed coverage with limited middle grounds [99]. Harsh and high-contrast lighting conditions often create artificial boundaries on viewed objects, altering appearances and reducing the accuracy of facial identification [99]. Over- and under-exposed footage may render an analysis impossible, based on multiple facial features being completely unrecognizable [21,106].

The capability of the camera is the primary factor in terms of low light or night-time conditions. Without the addition of directed lighting or dedicated "night-vision" cameras, CCTV systems must incorporate cameras that can record with infrared radiation (IR) and convert to visible light [107,108]. The accuracy of FFC has not been tested under IR conditions in our recent work and remains to be done in future studies. Most modern analogue and IP cameras are able to switch between day/night recording automatically and have IR LEDs built in to illuminate the target area. The range of the IR is generally limited to 20 m for midrange cameras on the market. This IR source of light could itself over-expose the subject dependent on SCD and other reflective materials present [99]. In addition, the IR footage is recorded in monochrome, and therefore,

includes the same limitations and challenges of traditional black and white CCTV footage in facial comparison [93], although in the current work, color was found to not be of much importance. Lens distortion effects and optical aberrations are more pronounced in IR cameras because of the longer wavelengths of IR [108]. Lastly, IR recording is subjected to image alterations and other artefacts based on converter quality and functioning [108]. Experimentally, IR recordings are difficult to conduct with subjects, as recording conditions need to be in low-to-zero light levels. There is a significant lack of research on IR CCTV recording in the context of facial identification and testing, and further validation of MA on a sample of IR surveillance data of comparable quality and conditions to the standard CCTV camera should be conducted.

As discussed above, placement and recording conditions of CCTV systems are crucial for reliable data capturing and use, especially in a forensic evidence context. This is inclusive of its installation in terms of network, software, and hardware. Many complications can arise as a result of these factors. Some examples experienced when attempting to develop the Wits Face Database [41,72] included inconsistent IP network connection and coverage, power outages, imminent weather problems, theft, and finally data loss, corruption, and tampering. Analogue CCTV systems for the most part do not provide remote video access and therefore require a physical storage and viewing location, limiting flexibility. These systems by default record at lower resolutions and require immediate local storage on a DVR device. This generally translates to a reduced amount of data loss and corruption compared to digital systems. Digital video can be recorded with varying rates of resolution, frame rate, and levels of compression [90]. The linkage of digital IP cameras to the internet allows for transmission of recorded footage for remote viewing, which requires high processing, storage, and data transmission capabilities. Digital video is, thus, more prone to occasional partial or complete data corruption or loss and is more perceptible to anti-forensic techniques, such as removing, hiding, and corrupting or wiping evidence from recorded footage [109–112]. In light of these threats, forensic readiness is needed in modern CCTV systems from both physical and cyber-attacks.

Little data exist describing the types and quantity of data loss incurred in CCTV systems globally and how this impacts surveillance and criminal investigations. Our studies [41] found approximately 21.2% loss of IP CCTV data and approximately 3.6% loss of analogue CCTV data during the establishment of the Wits Face Database (Table 1). CCTV data loss was noted in both IP and analogue cameras; however, the majority of corrupt or permanently lost data occurred with the digital IP camera systems. The CCTV systems utilized were an existing network at the university with no local storage and immediate transfer to a central server. During data transfer, any interruptions or fluctuations in local area network traffic or connectivity would result in data loss or irreparable corruption [41]. Studies utilizing existing CCTV systems and recordings are subjected to these types of data loss and corruption unless equipment is personally procured and installed. Data capture delays and reduced sample sizes are a considerable limitation when developing or expanding facial image databases.

The above discussed recommendations and primary limitations are generally not adhered to or considered, as is reflected in the actual data handed over to or available to law enforcement. Oftentimes, these data are of a subpar quality as a result of the numerous limitations as well as outdated camera systems [82]. Even with this subpar quality data and its limited utility, in a judiciary context, they may still successfully be implemented and should not be excluded until thoroughly reviewed first [113]. Thorough consideration of available evidence is in line with the ENFSI recommendations of triaging image data by their quality to ascertain fruitful use of FFC and efficient caseload management [114].

CCTV system installation and recording conditions are purpose driven and situationally applicable. They differ vastly to one another in terms of functionality, reliability, and environmental fit. System installation, hardware, and software need to be balanced in order to achieve the best result in terms of cost. Most systems are notably still lacking in applicability for facial comparison and are primarily disadvantaged not by installation and recording conditions but by image quality [20,82,89,90,92].

The poorer the derived image, whether it be from a photograph or CCTV footage, the lower the amount of extractable information. Image quality itself is a combination of multiple factors and related artefacts, with some of the relevant ones being resolution, pixelation, and noise. All of these are conditions that can vary notably across the various types of CCTV systems.

Analogue CCTV systems generally have lower resolutions and higher noise (grain) and often only record in monochrome. These cameras have been the global standard, and only in the last five years have we seen a large shift to internet protocol (IP) cameras [115,116]. The lower resolution leads to higher noise when attempting to enlarge the captured image for analysis and produces low clarity images [99,113].

The lack of color in most analogue recordings has a large impact in subsequent analysis, particularly in facial comparison [93]. Color plays an important role in face detection and recognition in humans, even when image quality is poor [117,118]. CCTV systems in general do not accurately capture color information from a scene [119] and have been deemed mostly unreliable in a forensic context [91,120]. Subject illumination as well as the color, orientation, and texture of objects are the primary variables dictating the accuracy of captured color information in CCTV [91]. When conducting MA using the FISWG facial feature list [47], color is the first component, and therefore, inaccurate image color data may lead to a decreased accuracy in performance. While color was easily disregarded in the majority of the analyses conducted in our studies, considering its contribution and consistency across CCTV recordings and photographs may be important for future studies.

More modern and commonly used internet protocol (IP) CCTV systems generally record full color at much higher resolutions with lower noise, as a result of high-spatial frequency blocking, overall leading to better extractable information for analysts [107]. Digital video is also a lot more flexible in recording and streaming quality compared with analogue in terms of video resolution, frame rate, and compression [90].

Poor quality CCTV recordings and extracted images have been shown to affect face matching ability in both novice and experts and leads to high overall false positive rates [82]. Image pixelation or spatial quantization, as a part of overall image quality, also drastically affects face matching ability [121]. Highly pixelated images can reduce face matching abilities by up to 50% in trained individuals when compared to a high-quality image sample [94,98]. In general, all forms of facial comparison accuracy will suffer when using low-resolution analogue CCTV images, even if image quality is good in other respects [48,81,93,122].

If we consider the SCD, the further away the subject, the greater the loss in detail in terms of representation of the face on the image. A minimum horizontal pixel count of 10–16 per face for a known face [121,123,124] and 20 pixels for an unknown face [92] is considered the bare minimum for successful identification in frontal view. Based on relative subject size on screen, Vitek et al. [125] recalculate Utochkin's [126] recommendations to 35 pixels for a known individual and 83 pixels for an unknown individual. If we are considering the effects of pixelation in a forensic setting, one needs to address the performance of matching accuracies and any form of potential enhancement, such as image blurring and reducing image size, when viewing [92]. Another important factor along the CCTV imaging chain not discussed here is that of the display fidelity and how the image is viewed on screen and the type of screen or monitor used [91].

Subject-to-camera distance can also result in facial distortion that alters facial proportions and shapes. While not exclusively investigated in the context of MA, previous work by Stephan and colleagues has looked at the SCD induced distortion and craniofacial superimposition [103,104,127]. Stephan [103] identified that discrepancy in camera-to-face, or skull, distances between photographs to be compared, presented with varying degrees of perspective distortion of facial features. At shorter distances, particularly below 6 m, the distortion was found to be more pronounced [103]. While distance-based perspective distortion is even more important in methods applying facial morphometry since at distances below 1 m a difference of 100 mm in SCD between the compared images can result in perspective distortion greater than 1% [103]. While perspective distortion would expectedly affect morphological comparison of facial features, the qualitative approach of MA and the large number of features being compared in each analysis would likely mitigate any small degrees of perspective distortion of compared images. However, further study into the effect of perspective distortion on qualitative assessment of facial features would be necessary.

The last, but important, limitation to consider in terms of image quality is video compression. As mentioned previously, digital video quality and corresponding file size can be made smaller in three ways—decreasing frame rate (e.g., 60 fps down to 5 fps); decreasing video resolution (e.g., Common image format (CIF) to Quarter CIF); and finally, by employing video compression [90,125,128]. Software video compression manipulates the spatial and temporal redundancy of moving frames in the form of CODECs, such as MPEG-4, Wavelet, H.265/HEVC, and JPEG [90,125,128]. Compression allows for large quantities of captured data to be stored in highly reduced sizes either temporarily or permanently but sacrifices image quality. Both distortion and artefacts occur when compression is introduced, hindering facial identification [90,125,129]. Keval and Sasse [90] found that the number of correct identifications of faces by untrained viewers decreased by 12–18% as MPEG-4 quality decreased and by 4–6% as Wavelet quality decreased (92–32 Kbps for both compression formats). They recommend a minimum of 52 Kbps video quality using MPEG-4 in order to achieve reliable and effective facial identification [90], albeit these results are for untrained practitioners and lower qualities would likely be reliable for trained FFC practitioners as well, perhaps not at the same magnitude. Vitek et al. [125] found correct identifications decreased from 88 to 48% as HEVC encoding quality decreased (30 kbps–15 kpbs) and they recommend 20 Kbps as a minimum threshold value. Compression employed in CCTV systems is lossy and, once performed during recording, cannot be removed or reversed. The types of distortion seen are pixelation, basis patterns, ringing, and blurring [99,130]. Recent advancements have been made improving FRT performance in light of compression artefacts; however, these artefacts remain a primary concern and drastically reduce accuracy and reliability [129]. An overview of the above-discussed various limiting factors of CCTV data in the application of MA and their specific effects in the process of facial comparison is presented in Table 2.

Table 2. Summary of CCTV systems' technical limitations in the application of morphological analysis.

General Limitations	Specific Limitations	Effects
Camera placement	• Camera height above ground [21,48,49,99,100] • Angle of incidence [21,48,49,100] • Subject-to-camera distance [103,104,127]	• Image composition affected—target size and screen/picture height [21,48,99] • Reduction of observable facial features [21,48,49,100] • Perspective distortion [103,104,127]

Table 2. *Cont.*

General Limitations	Specific Limitations	Effects
Camera specifications	• Analogue or digital [82,115,116] • Sensor size [99] • Pixel count [48,99] • Lens focal length [99]	• Reduced image quality [21,48,99] • Image distortion and artefacts [99]
Lighting conditions	• Ambient lighting [99,107,108] • Infrared vision [93,107,108]	• Loss of facial detail [48,49] • Shadows and overexposure form artificial boundaries and altered facial appearance [49,106] • Optical distortions [99]
Image quality	• Resolution [21,48] • Pixelation [48,92] • Noise/grain [99] • Video compression [90] • Color [48]	• Low clarity [99,103] • Reduced useable detail [21,48,82,90,92] • Face matching ability reduced [21,48,90,125,129]
Data loss and corruption	• Network infrastructure [41,72] • Software [99] • Hardware [41,99] • Imminent weather [41] • Power outages [41] • Compression rate [90] • Anti-forensic techniques [109–112]	• Inconsistent network connection and coverage—transfer corruption [41] • Partial or complete data loss [41] • Data tampering and removal [109–112]

In consideration of our results, only two particular CCTV camera specifications under limited conditions and installation variations were tested [21,48,49]. However, there is a large number of manufacturers that produce CCTV equipment with different specifications, requirements, and support. Testing the extent to which various market standard CCTV cameras can affect facial comparison would be an ideal goal to strive towards. However, before attempting such a level of fine-tuning of facial comparison practice and requirements, broader aspects should be investigated. These would include investigating the contributions of each of the various aspects that appeared to contribute to a decrease in MA performance, particularly in an attempt to determine empirical thresholds for suitable image quality across various specifications and not only image resolution. Therefore, the common factors described above that affect quality should be investigated. For instance, developing a thorough understanding of distance-related distortion effects on MA between images from CCTV cameras and photographs could generate awareness of which features are altered more notably, and hence, increase inaccuracy at unfavorable distances. This is an important consideration for future work due to the varied conditions most CCTV systems are installed under and tailored to. The alternative of comparing faces captured under the exact same conditions would likely be more effective; however, it may not be feasible or cost-effective outside of an experimental scenario. In addition, the time discrepancy between a first set of images from a CCTV recording and a recapture for analysis may introduce further limitations on the equipment and conditions of image capture (e.g., different lighting, damaged camera, etc.). In addition, studying the precise effect of camera angle of incidence on MA in isolation would also contribute to improving its application. Clear thresholds for determining the angle steepness that significantly inhibits facial comparison will aid in screening the utility of current image data and to guide future surveillance system installation planning. Incorporating the average head tilt in these investigations would further contribute to perfecting these standards beyond the experimental scenario.

Actual digital image quality and minimum resolution allowing for facial comparison to take place should also be investigated. Based on this study's conclusions with regard to low-resolution analogue CCTV, further investigations are needed in order to define a clear, quantifiable lower-end threshold that permits analysis. However, based on the actual accuracies and the approximate sizes of faces in each CCTV setting, it would appear that when a face is composed of approximately 18 × 26 pixels or less, such as the analogue CCTV setting employed in our study [48], FFC analysis would be severely compromised. This is a suggested preliminary lower threshold as despite conditions being mostly similar between standard CCTV and analogue CCTV, in terms of angle of incidence and SCD, faces in the standard CCTV were composed of approximately double the number of pixels (41 × 52 pixels) and a much higher accuracy and reliability were obtained [21]. This threshold remains well below Vitek et al.'s [125] and Utochkin's [126] recommendations (minimum of 83 pixels for unknown faces). Developing clear, experimentally tested lowest acceptable quality thresholds, particularly under different settings and conditions, will aid both the surveillance industry and the forensic analysts conducting analyses. A useful consideration for future studies investigating all aspects of image quality in facial comparison would be to use an image quality scoring system. An example of such a scale was presented by Schüler and Obertová [22]. Implementing this scale in conjunction with the FISWG feature list for MA could aid in identifying a threshold of confidence for the analysis process based on image quality.

Despite these uncertainties, from our earlier results, we recommend that CCTV system installations transition towards the use of high-definition cameras installed at eye-level heights. However, this would limit the cost-effectiveness of CCTV installations, as one camera would have a more limited field of view at the lower height [99]. As such, more cameras would need to be installed to cover areas previously covered by a single or pair of cameras [99]. Installing eye-level IP CCTV cameras would invariably place these systems at higher risk of vandalism and sabotage; however, the authors think this risk and increased cost are worthwhile in the context of facial comparison analyses, considering the significantly higher accuracy obtained when comparing faces recorded on these types of installations. Angle of incidence close to zero, allowing for more closely matching face views, in conjunction with high-resolution footage and the resulting quality of the facial image (at a minimum representation of a face being 41 × 52 pixels) are ideal for FFC application. While no clear benefit or shortfall of color recordings were isolated, based on the qualitative assessment of the analyses conducted, the authors would recommend the inclusion of color CCTV to allow for a wider range of feature list applications, such as the inclusion of color-based features in the FISWG feature list. However, we would also recommend the removal of color-based features as discrepancies in lighting are common between realistic recordings and ideal photographs captured for comparison. The resulting analysis of facial feature descriptors relying on color, or other factors that can vary easily and unknowingly, such as luminescence, in response to slight variations in lighting conditions should be reconsidered or removed from feature lists, as they were found either unreliable or unusable in most comparisons.

4.2. Feature List Usage, Disguises, and Training

Both the FISWG and the ENFSI recommend MA as the best practice for forensic facial identification [27,125]. In addition, FISWG advises against the use of photo-anthropometry for facial image comparison and recommends superimposition only to be utilized in conjunction with MA [27]. FISWG developed and made freely available an extensive facial feature list for use in MA [51,54]. This list includes 18 facial components, each with associated descriptors, as well as a nineteenth descriptive component for use with uncategorized features [51,54]. The FISWG feature list is also the most exhaustive list available, including over 130 facial component characteristics and over 290 characteristic descriptors [51].

The application of facial feature lists, such as the FISWG one, is different from previous feature-based comparison methods that involved facial feature classification schemes. These classification schemes were used as a way for an analyst to score each facial feature into categories based on descriptive qualities (e.g., pointed chin, broad nose bridge, etc.) [59]. The FISWG approach instead expects a facial analyst to subjectively describe the compared faces by providing an extensive list of features and descriptors to use in order to make statements based on similarities and dissimilarities [27,51]. This descriptive approach is preferred as classification schemes are viewed as prone to high inter-observer error [55,59,109]. In addition, for classification schemes to be effective, they need to be tailored to specific populations, which has only been considered by two studies to date [58,127]. Population homogeneity, however, can be problematic for classification schemes, since high prevalence of a feature classification in a given population could result in an overlapping score, leading to erroneous false positive matches [60]. On the other hand, classification schemes may be too restrictive and make scoring near impossible under certain circumstances [26]. In this series of studies, the FISWG feature list was found to greatly aid both in the training of the analyst as well as during the analysis process to achieve mostly high accuracies and good reliability levels with the exception of the lowest quality of CCTV recordings. The feature list was also found to be applicable to African male faces due to its descriptive nature, as opposed to population-specific classification schemes. Certain descriptors were found cumbersome to utilize; for example, as mentioned above, skin color and luminance were often ignored due to a mismatch, despite confirmation that two faces were indeed the same. A revision of some of these descriptors would be required to optimize the analysis process and applicability of the FISWG feature list to a broader number of settings and CCTV conditions.

In addition, while the importance of a feature list in MA is undeniable, combining a systematic approach and a feature list must involve the option to discard potential dissimilarities when it is justifiable to do so. This is possible with large feature lists such as the FISWG one, which allows for exclusion of questionable or hard to analyze features. A smaller feature list be employed would compromise the exclusion of dissimilar features that could be justified as dissimilar due to image conditions, leading to the false exclusion of a positive face match due to features varying under the different image conditions. To this end, a threshold of the number of minimum features required to conduct an analysis should be investigated as neither the feature list [47] nor the concluding statements [17] provided one.

A further consideration to improve the applicability of the FISWG and any other feature lists would be to develop specific criteria to be applied for comparisons under different disguised or obstructed faces. Once established, these criteria could be included in analyst training to prioritize features by type of disguise. This approach would be applicable in settings where facial features may not be visible due to data loss or any other physical obstructions. This could prove particularly useful as the forms of "acceptable disguises" change throughout time—for example the use of face masks currently due to the spread of COVID-19. Face masks, which can vary in shape, size, and the resulting proportion of the face covered, have been shown to reduce automated facial recognition performance by 5 to 50% depending on the specific algorithm and extent of the face covered [76]. The deleterious effect in performance seemed to vary based on the color and shape of the masks as well [76]. It would, hence, be crucial to consider face masks in further tests of MA under disguised conditions as their impact on human observer-based facial comparison has not yet been considered.

While the current studies validated MA when the FISWG guidelines [27,47] were used, an array of further studies is required. To continue this work, the Wits Face Database will need to be updated and expanded to include many more possible permutations of analysis, including female individuals, cosmetic make-up treatments, and face mask disguises. Future studies should also attempt to quantify the acceptable loss of facial feature information in order to successfully compare faces across a multitude of possible situations. Once that goal is achieved, better guidelines and practice frameworks can be created for legal procedures involving FFC by MA. Our recommended method of conducting MA is shown in Figure 3. This approach outlines the stepwise process of applying the recommended image quality triage by the ENFSI [114] to FISWG's ACE-V application [27] of the FISWG feature list [47] and then subdivided by intended use in a research or judicial context. Based on the application context, different approaches are used for verification; in addition, research use of MA would require further statistical analyses (Figure 3).

Based on the outcomes of our group's studies, expanding current training programs and developing new ones to increase the competence of facial comparison experts will be crucial for consistent and reliable application of FFC. The application of a feature list and an ACE-V approach by members of the public in forensic facial comparisons is not sufficient to achieve expertise. Training experts with the explicit role to conduct FFC analyses, with the use of the FISWG feature list and an ACE-V approach, is of utmost importance in a judicial context [131]. The role of expertise is particularly relevant in FFC, since unfamiliar face matching is considered complex and unreliable on all accounts [97,122]. Experts, in fact, perform notably better than members of the public [42], even when image quality was taken into consideration [132]. This expertise undoubtably arises from training in the nuances of faces, such as facial expressions and ageing changes, and acceptable anatomical variations and image-based variations between faces, beyond just the inclusion of the use of a feature list. The need for adequate training of all FFC practitioners is vital to the good standing of the practice and its admissibility in a legal context. Particularly when considering that certain countries may experience a shortage of expertise and heavy caseloads, such as South Africa, where only 30 trained specialists in the entirety of the national police force are trained to conduct FFC and testify in court to defend their conclusions [17].

The FISWG has put forward a document describing guidelines for training and expertise requirements of FFC analysts and trainers [133]. While these guidelines are crucial to the development of training courses, to the authors' knowledge, no formal standardized training or certification platforms exist for FFC [17]. A recent study on the performance of informal training courses on facial comparison suggested that there are large discrepancies between courses in improvement of facial examiner expertise [134]. We hope our recommended stepwise process to the applications of MA (Figure 3) will aid in streamlining both MA training and application. Recently, members of our research group proposed an outline for a training course with a three-tiered approach offered to the police force [17]; it inadvertently follows most of the proposed guidelines from FISWG. The first tier of training involves developing basic background knowledge of facial anatomy, evidence evaluation, image science, facial recognition psychology, and court proceedings, among other topics [17]. The second tier involves training in detailed MA using the FISWG standards and developing court-ready reports and charts [17]. The third tier is a national specific tier that involves advanced training in court proceedings and evidence presentation as well as troubleshooting from past casework in order to also train experienced peer reviewers [17], who are vital to the ACE-V application of MA. While this approach to training has not been experimentally tested, the trained police members have found success in their roles as facial examiners.

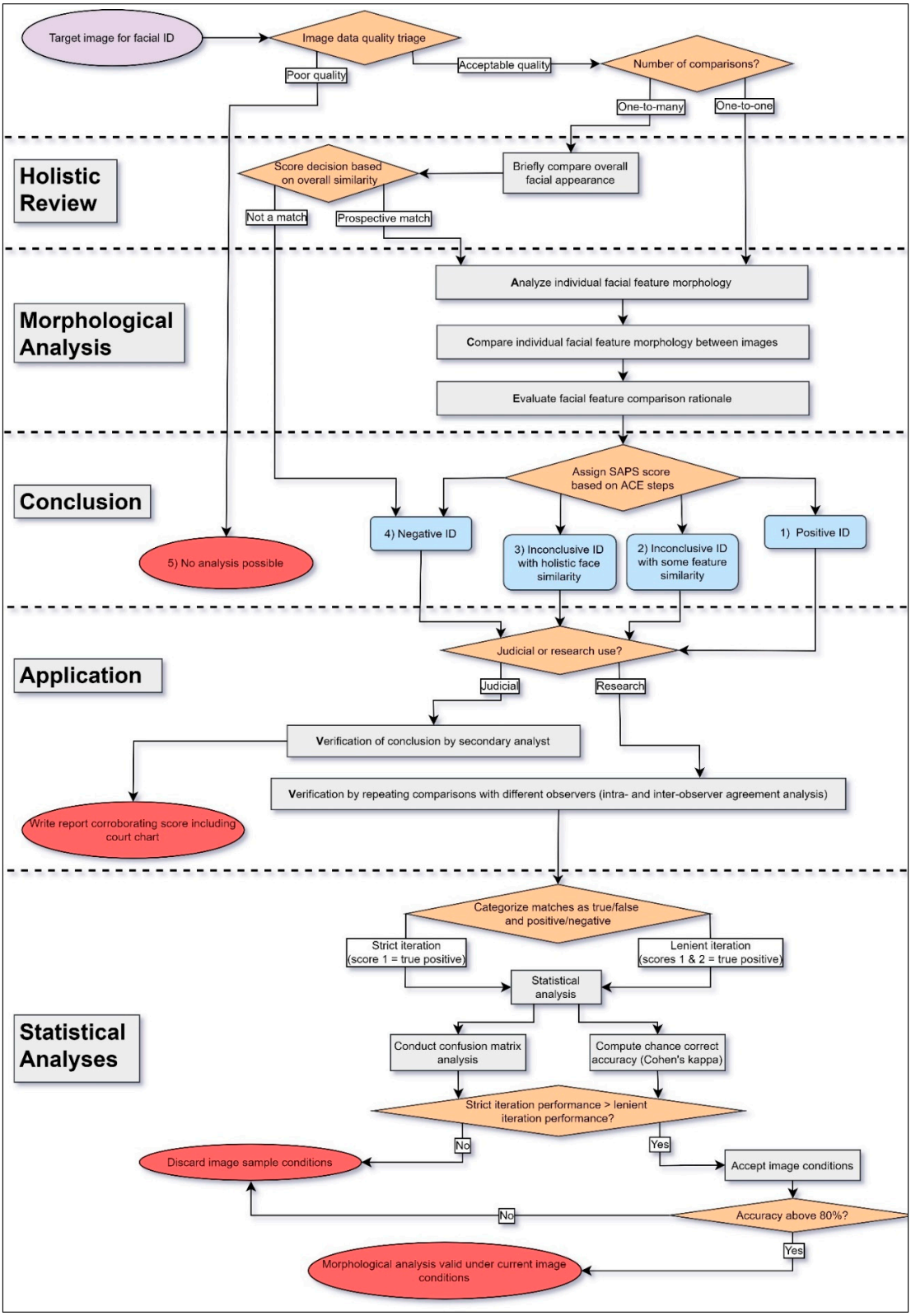


Figure 3. Flow diagram of the recommended morphological analysis process. This approach to morphological analysis uses an ACE-V method in conjunction with the FISWG feature list [47], with the inclusion of the ENFSI's image quality triaging [114] and the use of the South African Police Services (SAPS) scoring criteria [17] as adapted for research application [21]. Statistical analyses for research use are also recommended based on our recent work [48] to allow for more detailed result interpretation and comparison among future studies.

5. Conclusions

The outcomes and recommendations arising from these studies should be considered under the limitations of the investigative approach deployed. These studies attempted to simulate real-world conditions, with its array of limitations, in a select number of scenarios. These scenarios included ideal comparable photographic images, standard digital CCTV, eye-level installation digital CCTV, standard monochrome analogue CCTV, and two common disguises—sunglasses and brimmed caps. The varied circumstances of facial data were pre-set under certain conditions to attempt some level of standardization required for experimentation. In doing so, although realistic, the conditions were limited to the major questions broadly investigated by each study [21,48,49]. Although these broad categories were considered as realistic examples of CCTV image quality and conditions, there are multiple factors that influence the quality of an image for facial identification. Image quality relies on more than just equipment resolution capacity; it involves lighting conditions, angle of incidence, SCD, distortions, color, visibility of features, and more. In testing the specific conditions outlined in each of the above studies, controlling for or identifying which of the multiple limiting factors contributed to the poor performance of MA would be impossible. Even when within likely tolerable degrees, these limiting factors cannot be isolated from one another in certain circumstances. However, with this baseline of conditions and considerations, future studies can be tailored to the specific limitations that CCTV imposes on FFC in a highly controlled setting to determine the exact contribution of each of these limiting factors to the accuracy of MA.

With these concerns and limitations clearly stated, future studies should be focused to target specific limiting factors individually in order to develop a clear threshold for image data to be usable for facial comparison. While other approaches to facial identification as a whole may also be gaining popularity, such as the increasing performance of automated systems [31,75,76] and the deployment of super-recognizers [135,136], continued research in forensic facial comparison by MA is crucial as the most universally applicable and reliable method. The importance of MA-based FFC is especially noteworthy in law enforcement applications, where the majority of available image data is of low to poor quality [113]. As such, the authors strongly advise that trained human observer-based MA, using the FISWG feature list [47] and an ACE-V approach [27], should remain the principal method of facial comparison for identification purposes, as recommended by both the FISWG and ENFSI [27,114].

Author Contributions: Conceptualization, N.B. (Nicholas Bacci), M.S. and N.B. (Nanette Briers); methodology, N.B. (Nicholas Bacci) and J.G.D.; formal analysis, N.B. (Nicholas Bacci); investigation, N.B. (Nicholas Bacci); visualization, N.B. (Nicholas Bacci) and J.G.D.; resources, M.S. and N.B. (Nanette Briers); data curation, N.B. (Nicholas Bacci) and J.G.D.; writing—original draft preparation, N.B. (Nicholas Bacci), J.G.D. and N.B. (Nanette Briers); writing—review and editing, N.B. (Nicholas Bacci), J.G.D., M.S. and N.B. (Nanette Briers); supervision, M.S. and N.B. (Nanette Briers); project administration, N.B. (Nicholas Bacci); funding acquisition, N.B. (Nicholas Bacci), M.S. and N.B. (Nanette Briers). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the South African National Research Foundation (NRF), DAAD-NRF Bursary, and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand awarded to N. Bacci (Grant No.: 11858) and NRF funds awarded to N. Briers as part of the Improving Methodologies and Practices in Cranio-facial Identification (Grant No.: CSUR160425163022; UID: 106031). Any opinions, findings, and conclusions or recommendations expressed in this study are those of the authors, and therefore, the NRF and University of the Witwatersrand, Johannesburg, do not accept any liability in regard thereto. The funders were not involved in the study design, data collection, analysis, interpretation, or writing of the manuscript or submission. The APC was funded by N. Briers' and M. Steyn's Research Incentive (RINC) funds.

Institutional Review Board Statement: The core studies [21,48,49] discussed in the present review paper were conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg. The ethics approval number is M171026, and approval was granted on 13 March 2018.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the core studies [21,48,49] discussed in the present review paper. Written informed consent has been obtained from the corresponding author to publish his facial images as part of this paper.

Data Availability Statement: The facial image data used in Figures 1 and 2 for the current paper are of the corresponding author and publicly available as part of the sample images of the Wits Face Database [41], reproducible under an open access license distributed under the terms of the Creative Commons Attribution License. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. These images can be accessed as part of the Wits Face Database data note, via the supplementary material for the Wits Face Database [41].

Acknowledgments: Thanks are due to the support provided from the South African National Research Foundation (NRF), DAAD-NRF Bursary, and the J.J.J. Smieszek Fellowship from the School of Anatomical Sciences, University of the Witwatersrand. DAAD-NRF and J.J.J. Smieszek Fellowship funds were awarded to N. Bacci (Grant No.: 11858) and NRF funds were awarded to N. Briers as part of the Improving Methodologies and Practices in Cranio-facial Identification (Grant No.: CSUR160425163022; UID: 106031). Any opinions, findings, and conclusions or recommendations expressed in this study are those of the authors, and therefore, the NRF and University of the Witwatersrand, Johannesburg, do not accept any liability in regard thereto. The funders were not involved in study design, data collection, analysis, interpretation, or writing of the manuscript or its submission. Acknowledgement is also owed to all the participants who agreed to be photographed and recorded for the development of the Wits Face Database. Thanks are due to Tamara Lottering for her assistance in composing the face pools and Gideon LeRoux for his assistance in the capturing and extracting of the CCTV recordings from the University security systems. Gratitude is also owed to Tobias Houlton for his aid in study validation. Lastly, thanks are due to the volunteers who aided in participant recruitment for the Wits Face Database: Jesse Fredericks, Kiveshen Pillay, Rethabile Masiu, Sameerah Sallie, Daniel Munesamy, Laurette Joubert, Jordan Swiegers, Betty Mkabela, Johannes P. Meyer, Amy Spies, Natasha Loubser, Nicole Virgili, Dan-Joel Lukumbi, Tamara Lottering, Mathabatha Ntjie, Claudia Landsman, Raheema Dalika, Merete Goosen, Stephanie Souris, Rabelani Negota, Mahlatse Mahasha, and Jessica Manavhela.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Jäger, J. Photography: A means of surveillance? Judicial photography, 1850 to 1900. *Crime Hist. Sociétés* **2001**, *5*, 27–51. [CrossRef]
2. Bertillon, A.; McClaughry, R.W. *Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification*; McClaughry, R.W., Ed.; The Werner Company: Chicago, IL, USA, 1896.
3. Faigman, D.L. Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science. *Hastings Law J.* **2008**, *59*, 979–1000.
4. Mokwena, R.J. *The Value of Photography in the Investigation of Crime Scenes*; University of South Africa: Pretoria, South Africa, 2012.
5. Bell, A. Crime scene photography in England, 1895–1960. *J. Br. Stud.* **2018**, *57*, 53–78. [CrossRef]
6. Lindegaard, M.R.; Bernasco, W. Lessons Learned from Crime Caught on Camera. *J. Res. Crime Delinq.* **2018**, *55*, 155–186. [CrossRef] [PubMed]
7. Norris, C.; McCahill, M.; Wood, D. The Growth of CCTV: A global perspective on the international diffusion of video surveillance in publicly accessible space. *Surveill. Soc.* **2002**, *2*, 110–135. [CrossRef]
8. Piza, E.L.; Welsh, B.C.; Farrington, D.P.; Thomas, A.L. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminol. Public Policy* **2019**, *18*, 135–159. [CrossRef]
9. Jain, A.K.; Klare, B.; Park, U. Face Matching and Retrieval in Forensics Applications. *IEEE Multimed.* **2012**, *19*, 20. [CrossRef]
10. Moyo, S. *Evaluating the Use of CCTV Surveillance Systems for Crime Cotnrol and Prevention: Selected Case Studies from Johannesburg and Tshwane, Gauteng*; University of South Africa: Pretoria, South Africa, 2019.
11. Goold, B.; Loader, I.; Thumala, A. The banality of security: The curious case of surveillance cameras. *Br. J. Criminol.* **2013**, *53*, 977–996. [CrossRef]
12. Duncan, J. How CCTV surveillance poses a threat to privacy in South Africa. *Conversation* **2018**, 1–3. Available online: https://theconversation.com/how-cctv-surveillance-poses-a-threat-to-privacy-in-south-africa-97418 (accessed on 30 October 2021).
13. Gill, M.; Spriggs, A. Assessing the Impact of CCTV: Home Office Research Study 292. 2005. Available online: https://techfak.uni-bielefeld.de/~{}iluetkeb/2006/surveillance/paper/social_effect/CCTV_report.pdf (accessed on 25 November 2021).

14. Welsh, B.C.; Farrington, D.P. Effects of Closed Circuit Television Surveillance on Crime. *Campbell Syst. Rev.* **2008**, *4*, 1–73. [CrossRef]
15. Ashby, M.P.J. The Value of CCTV Surveillance Cameras as an Investigative Tool: An Empirical Analysis. *Eur. J. Crim. Policy Res.* **2017**, *23*, 441–459. [CrossRef]
16. Kleinberg, K.F.; Siebert, J.P. A study of quantitative comparisons of photographs and video images based on landmark derived feature vectors. *Forensic Sci. Int.* **2012**, *219*, 248–258. [CrossRef]
17. Steyn, M.; Pretorius, M.; Briers, N.; Bacci, N.; Johnson, A.; Houlton, T.M.R. Forensic facial comparison in South Africa: State of the science. *Forensic Sci. Int.* **2018**, *287*, 190–194. [CrossRef]
18. Jackson, A. The Admissibility of Identification Evidence Made on the Basis of Recognition from Photographs Taken at a Crime Scene. *J. Crim. Law* **2016**, *80*, 234–236. [CrossRef]
19. Houlton, T.M.R.; Steyn, M. Finding Makhubu: A morphological forensic facial comparison. *Forensic Sci. Int.* **2018**, *285*, 13–20. [CrossRef]
20. Stephan, C.N.; Caple, J.M.; Guyomarc'h, P.; Claes, P. An overview of the latest developments in facial imaging. *Forensic Sci. Res.* **2019**, *4*, 10–28. [CrossRef] [PubMed]
21. Bacci, N.; Houlton, T.M.R.; Briers, N.; Steyn, M. Validation of forensic facial comparison by morphological analysis in photographic and CCTV samples. *Int. J. Legal Med.* **2021**, *135*, 1965–1981. [CrossRef] [PubMed]
22. Schüler, G.; Obertová, Z. Visual identification of persons: Facial image comparison and morphological comparative analysis. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, A., Cattaneo, C., Eds.; Elsevier Academic Press: London, UK, 2020; pp. 313–330.
23. Behrman, B.W.; Davey, S.L. Eyewitness identification in actual criminal cases: An archival analysis. *Law Hum. Behav.* **2001**, *25*, 475–491. [CrossRef]
24. Boyce, M.A.; Lindsay, D.S.; Brimacombe, C.A.E. Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators. *Law Hum. Behav.* **2008**, *32*, 439–453. [CrossRef] [PubMed]
25. Davis, J.P.; Valentine, T.; Wilkinson, C. Facial image comparison. In *Craniofacial Identification*; Wilkinson, C., Rynn, C., Eds.; Cambridge University Press: New York, NY, USA, 2012; pp. 136–153. ISBN 9781139049566.
26. Valentine, T.; Davis, J.P. *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV*; John Wiley & Sons, Ltd.: Chichester, UK, 2015; ISBN 9781118469538.
27. Facial Identification Scientific Working Group. Facial Comparison Overview and Methodology Guidelines. 2019. Available online: https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf (accessed on 25 November 2021).
28. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [CrossRef]
29. Akhtar, Z.; Rattani, A. A Face in any Form: New Challenges and Opportunities for Face Recognition Technology. *Computer* **2017**, *50*, 80–90. [CrossRef]
30. Lai, X.; Patrick Rau, P.L. Has facial recognition technology been misused? A user perception model of facial recognition scenarios. *Comput. Hum. Behav.* **2021**, *124*, 106894. [CrossRef]
31. Kaur, P.; Krishan, K.; Sharma, S.K.; Kanchan, T. Facial-recognition algorithms: A literature review. *Med. Sci. Law* **2020**, *60*, 131–139. [CrossRef] [PubMed]
32. Grother, P.; Ngan, M.; Hanaoka, K. *Face Recognition Vendor Test (FRVT) Part 2: Identification*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2019.
33. Dodd, V. UK Police Use of Facial Recognition Technology a Failure, Says Report. *The Guardian*, 15 May 2018.
34. Grother, P.; Ngan, M.; Hanaoka, K. *Face Recognition Vendor Test Part 3: Demographic Effects*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, December 2019.
35. White, D.; Jonathon Phillips, P.; Hahn, C.A.; Hill, M.; O'Toole, A.J. Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20151292. [CrossRef]
36. Davis, J.P.; Valentine, T.; Davis, R.E. Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Sci. Int.* **2010**, *200*, 165–176. [CrossRef] [PubMed]
37. Urbanová, P. Performance of distance-based matching algorithms in 3D facial identification. *Egypt. J. Forensic Sci.* **2016**, *6*, 135–151. [CrossRef]
38. Press Association. Welsh Police Wrongly Identify Thousands as Potential Criminals. *The Guardian*, 5 May 2018.
39. Mordor Intelligence. Global Facial Recognition Market (2021–2026). 2020. Available online: https://www.mordorintelligence.com/industry-reports/facial-recognition-market (accessed on 30 October 2021).
40. Spaun, N.A. Facial comparisons by subject matter experts: Their role in biometrics and their training. *Int. Conf. Biom.* **2009**, *5558*, 161–168. [CrossRef]
41. Bacci, N.; Davimes, J.; Steyn, M.; Briers, N. Development of the Wits Face Database: An African database of high-resolution facial photographs and multimodal closed-circuit television (CCTV) recordings. *F1000Research* **2021**, *10*, 131. [CrossRef]
42. Wilkinson, C.; Evans, R. Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Sci. Justice* **2009**, *49*, 191–196. [CrossRef]

43. Valentine, T.; Davis, J.P. *Forensic Facial Identification*; John Wiley & Sons: Hoboken, NJ, USA, 2015; ISBN 9781118469118.
44. White, D.; Dunn, J.D.; Schmid, A.C.; Kemp, R.I. Error Rates in Users of Automatic Face Recognition Software. *PLoS ONE* **2015**, *10*, e0139827. [CrossRef]
45. Speckeis, C. Can ACE-V be validated? *J. Forensic Identif.* **2011**, *61*, 201–209.
46. Norris, C. The success of failure. Accounting for the global growth of CCTV. In *Routledge Handbook of Surveillance Studies*; Ball, K., Haggerty, K.D., Lyon, D., Eds.; Routledge: London, UK; New York, NY, USA, 2012; pp. 251–258.
47. Facial Identification Scientific Working Group. Facial Image Comparison Feature List for Morphological Analysis. 2018. Available online: https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf (accessed on 25 November 2021).
48. Bacci, N.; Steyn, M.; Briers, N. Performance of forensic facial comparison by morphological analysis across optimal and suboptimal CCTV settings. *Sci. Justice* **2021**, *61*, 743–754. [CrossRef]
49. Bacci, N.; Briers, N.; Steyn, M. Assessing the effect of facial disguises on forensic facial comparison by morphological analysis. *J. Forensic Sci.* **2021**, *66*, 1220–1233. [CrossRef]
50. Martinez, A.M.; Benavente, R. *The AR Face Database CVC Technical Report #24*; Universitat Autònoma de Barcelona: Bellaterra, Spain, 1998; Volume 24.
51. Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1090–1104. [CrossRef]
52. Sim, T.; Baker, S.; Bsat, M. The CMU Pose, Illumination, and Expression (PIE) database. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1615–1618.
53. Calvo, M.G.; Lundqvist, D. Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behav. Res. Methods* **2008**, *40*, 109–115. [CrossRef] [PubMed]
54. Gao, W.; Cao, B.; Shan, S.; Chen, X.; Zhou, D.; Zhang, X.; Zhao, D. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Trans. Syst. Man Cybern.* **2008**, *38*, 149–161.
55. Milborrow, S.; Morkel, J.; Nicolls, F. The MUCT Landmarked Face Database. *Pattern Recognit. Assoc. S. Afr.* **2008**. Available online: http://www.milbo.org/muct/ (accessed on 25 November 2021).
56. Tottenham, N.; Tanaka, J.W.; Leon, A.C.; McCarry, T.; Nurse, M.; Hare, T.A.; Marcus, D.J.; Westerlund, A.; Casey, B.J.; Nelson, C. The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res.* **2009**, *168*, 242–249. [CrossRef]
57. Grgic, M.; Delac, K.; Grgic, S. SCface—Surveillance cameras face database. *Multimed. Tools Appl.* **2011**, *51*, 863–879. [CrossRef]
58. Kostinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; IEEE: Barcelona, Spain, 2011; pp. 2144–2151.
59. Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 Faces In-The-Wild Challenge: Database and results. *Image Vis. Comput.* **2015**, *47*, 3–18. [CrossRef]
60. Karam, L.J.; Zhu, T. Quality labeled faces in the wild (QLFW): A database for studying face recognition in real-world environments. *Hum. Vis. Electron. Imaging XX* **2015**, *9394*, 93940. [CrossRef]
61. Ma, D.S.; Correll, J.; Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **2015**, *47*, 1122–1135. [CrossRef]
62. Belhumeur, P.N.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [CrossRef]
63. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]
64. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; University of Massachusetts: Amherst, MA, USA, 2007; Volume October, pp. 1–11.
65. Gross, R. Face Databases. In *Handbook Face Recognition*; Springer: New York, NY, USA, 2005; pp. 301–327. [CrossRef]
66. Shapiro, P.N.; Penrod, S. Meta-analysis of facial identification studies. *Psychol. Bull.* **1986**, *100*, 139–156. [CrossRef]
67. Bothwell, R.K.; Brigham, J.C.; Malpass, R.S. Cross-Racial Identification. *Pers. Soc. Psychol. Bull.* **1989**, *15*, 19–25. [CrossRef]
68. Anthony, T.; Copper, C.; Mullen, B. Cross-Racial Facial Identification: A Social Cognitive Integration. *Pers. Soc. Psychol. Bull.* **1992**, *18*, 296–301. [CrossRef]
69. Meissner, C.A.; Brigham, J.C. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol. Public Policy Law* **2001**, *7*, 3–35. [CrossRef]
70. Chiroro, P.M.; Tredoux, C.G.; Radaelli, S.; Meissner, C.A. Recognizing faces across continents: The effect of within-race variations on the own-race bias in face recognition. *Psychon. Bull. Rev.* **2008**, *15*, 1089–1092. [CrossRef] [PubMed]
71. Kokje, E.; Bindemann, M.; Megreya, A.M. Cross-race correlations in the abilities to match unfamiliar faces. *Acta Psychol.* **2018**, *185*, 13–21. [CrossRef]
72. Bacci, N.; Davimes, J.; Steyn, M.; Briers, N. Wits Face Database. Available online: https://hdl.handle.net/10539/29924 (accessed on 30 October 2021).
73. Bromby, M. CCTV and Expert Evidence: Addressing the Reliability of New Sciences. *Archbold News*, 2 November 2006; 6–9.
74. Kuhn, M. Caret: Classification and Regression Training R Package. 2020. Available online: https://cran.r-project.org/web/packages/caret/index.html (accessed on 30 October 2021).

75. Ngan, M.; Grother, P.; Hanaoka, K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2020.
76. Ngan, M.; Grother, P.; Hanaoka, K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face Recognition Accuracy with Face Masks Using Post-COVID-19 Algorithms*; US Department of Commerce, National Institute of Standards & Technology: Gaithersburg, MD, USA, 2020.
77. Davis, J.P.; Valentine, T. CCTV on trial: Matching video images with the defendant in the dock. *Appl. Cogn. Psychol.* **2009**, *23*, 482–505. [CrossRef]
78. Bruce, V.; Henderson, Z.; Greenwood, K.; Hancock, P.J.B.; Burton, A.M.; Miller, P. Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* **1999**, *5*, 339–360. [CrossRef]
79. Kemp, R.I.; Caon, A.; Howard, M.; Brooks, K.R. Improving Unfamiliar Face Matching by Masking the External Facial Features. *Appl. Cogn. Psychol.* **2016**, *30*, 622–627. [CrossRef]
80. Terry, R.L. How wearing eyeglasses affects facial recognition. *Curr. Psychol.* **1993**, *12*, 151–162. [CrossRef]
81. Henderson, Z.; Bruce, V.; Burton, A.M. Matching the faces of robbers captured on video. *Appl. Cogn. Psychol.* **2001**, *15*, 445–464. [CrossRef]
82. Lee, W.-L.; Wilkinson, C.; Memon, A.; Houston, K. Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: An evaluation of the effect of training on facial identification ability. *Axis Online J. CAHId* **2009**, *1*, 19–28.
83. Righi, G.; Peissig, J.J.; Tarr, M.J. Recognizing disguised faces. *Vis. Cogn.* **2012**, *20*, 143–169. [CrossRef]
84. Towler, A.; White, D.; Kemp, R.I. Evaluating the feature comparison strategy for forensic face identification. *J. Exp. Psychol. Appl.* **2017**, *23*, 47–58. [CrossRef] [PubMed]
85. Megreya, A.M.; Bindemann, M. Feature instructions improve face-matching accuracy. *PLoS ONE* **2018**, *13*, e0193455. [CrossRef] [PubMed]
86. Megreya, A.M. Feature-by-feature comparison and holistic processing in unfamiliar face matching. *PeerJ* **2018**, *6*, e4437. [CrossRef]
87. Goold, B.J. Open to all? Regulating open street CCTV and the case for "symmetrical surveillance". *Crim. Justice Ethics* **2006**, *25*, 3–17. [CrossRef]
88. Welsh, B.C.; Farrington, D.P. Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Q.* **2009**, *26*, 716–745. [CrossRef]
89. Kleinberg, K.F.; Vanezis, P.; Burton, A.M. Failure of anthropometry as a facial identification technique using high-quality photographs. *J. Forensic Sci.* **2007**, *52*, 779–783. [CrossRef] [PubMed]
90. Keval, H.U.; Sasse, M.A. Can we ID from CCTV? Image quality in digital CCTV and face identification performance. *Mob. Multimedia Image Process. Secur. Appl.* **2008**, *6982*, 69820. [CrossRef]
91. Smith, R.A.; MacLennan-Brown, K.; Tighe, J.F.; Cohen, N.; Triantaphillidou, S.; MacDonald, L.W. Colour analysis and verification of CCTV images under different lighting conditions. *Image Qual. Syst. Perform. V* **2008**, *6808*, 68080. [CrossRef]
92. Bindemann, M.; Attard, J.; Leach, A.; Johnston, R.A. The effect of image pixelation on unfamiliar-face matching. *Appl. Cogn. Psychol.* **2013**, *27*, 707–717. [CrossRef]
93. Burton, A.M.; Wilson, S.; Cowan, M.; Bruce, V. Face recognition in poor-quality video: Evidence from security surveillance. *Psychol. Sci.* **1999**, *10*, 243–248. [CrossRef]
94. Ritchie, K.L.; White, D.; Kramer, R.S.S.; Noyes, E.; Jenkins, R.; Burton, A.M. Enhancing CCTV: Averages improve face identification from poor-quality images. *Appl. Cogn. Psychol.* **2018**, *32*, 671–680. [CrossRef]
95. Fysh, M.C.; Bindemann, M. The Kent Face Matching Test. *Br. J. Psychol.* **2018**, *109*, 219–231. [CrossRef]
96. Kramer, R.S.S.; Mohamed, S.; Hardy, S.C. Unfamiliar Face Matching With Driving Licence and Passport Photographs. *Perception* **2019**, *48*, 175–184. [CrossRef] [PubMed]
97. Burton, A.M.; White, D.; McNeill, A. The Glasgow Face Matching Test. *Behav. Res. Methods* **2010**, *42*, 286–291. [CrossRef]
98. Bindemann, M.; Attard, J.; Johnston, R.A. Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychol.* **2014**, *1*, 986903. [CrossRef]
99. Damjanovski, V. *CCTV from Light to Pixels*, 3rd ed.; Elsevier: Oxford, UK, 2014; ISBN 9780124045576.
100. Ward, D. Testing Camera Height vs. Image Quality; Pennsylvania, USA. 2013. Available online: https://ipvm.com/reports/testing-camera-height (accessed on 25 November 2021).
101. Mileva, M.; Burton, A.M. Face search in CCTV surveillance. *Cogn. Res. Princ. Implic.* **2019**, *4*, 1–21. [CrossRef]
102. Abudarham, N.; Yovel, G. Reverse engineering the face space: Discovering the critical features for face identification. *J. Vis.* **2016**, *16*, 40. [CrossRef] [PubMed]
103. Stephan, C.N. Perspective distortion in craniofacial superimposition: Logarithmic decay curves mapped mathematically and by practical experiment. *Forensic Sci. Int.* **2015**, *257*, 520.e1–520.e8. [CrossRef]
104. Stephan, C.N.; Armstrong, B. Scientific estimation of the subject-to-camera distance from facial photographs for craniofacial superimposition. *Forensic Sci. Int. Rep.* **2021**, *4*, 100238. [CrossRef]
105. British Security Industry. Video Surveillance Systems for Use in Security Applications: System Requirements—General. 2014. Available online: https://www.bsia.co.uk/publications/video-surveillance/ (accessed on 30 October 2021).
106. Cohen, N.; Gattuso, J.; MacLennan-Brown, K. *CCTV Operational Requirements Manual*; Cohen, N., Gattuso, J., MacLennan-Brown, K., Eds.; Home Office Scientific Development Branch: Sandridge, UK, 2009; ISBN 9781847269027.

107. Surette, R. The thinking eye: Pros and cons of second generation CCTV surveillance systems. *Policing* **2005**, *28*, 152–173. [CrossRef]
108. Kruegle, H. *CCTV Surveillance: Analog and Digital Video Practices and Technology*, 2nd ed.; Elsevier Butterworth–Heinemann: Burlington, MA, USA, 2007; ISBN 9780750677684.
109. Blunden, B. Anti-Forensics: The Rootkit Connection. In Proceedings of the Black Hat USA 2009, Las Vegas, NV, USA, 25–30 July 2009; pp. 1–44.
110. D'Orazio, C.; Ariffin, A.; Choo, K.-K.R. IOS Anti-Forensics: How Can We Securely Conceal, Delete and Insert Data? In Proceedings of the 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 13 October 2013; pp. 6–9. Available online: https://ssrn.com/abstract=2339819 (accessed on 30 October 2021).
111. Kissel, R.; Regenscheid, A.; Scholl, M.; Stine, K. *Guidelines for Media Sanitization*; US Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2014; Volume 800.
112. Ariffin, A.; Choo, K.K.; Yunos, Z. Forensic readiness: A case study on digital CCTV systems antiforensics. In *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*; Choo, K.-K.R., Dehghantanha, A., Eds.; Syngress: London, UK, 2017; pp. 147–162.
113. Gibelli, D.; Obertová, Z.; Ritz-Timme, S.; Gabriel, P.; Arent, T.; Ratnayake, M.; De Angelis, D.; Cattaneo, C. The identification of living persons on images: A literature review. *Leg. Med.* **2016**, *19*, 52–60. [CrossRef]
114. ENFSI. *Best Practice Manual for Facial Image Comparison*; ENFSI: Wiesbaden, Germany, 2018; Volume 1, Available online: https://enfsi.eu/wp-content/uploads/2017/06/ENFSI-BPM-DI-01.pdf (accessed on 30 October 2021).
115. Smith, S. CCTV Market Outlook 2017. *Cision PR Newswire*, 15 May 2014.
116. Wood, L. CCTV Cameras—Worldwide Market Outlook Report 2018–2026: Dome Cameras Dominate. *Businesswire* **2018**. Available online: https://www.businesswire.com/news/home/20180913005519/en/CCTV-Cameras---Worldwide-Market-Outlook-Report-2018-2026-Dome-Cameras-Dominate---ResearchAndMarkets.com (accessed on 30 October 2021).
117. Yip, A.W.; Sinha, P. Contribution of color to face recognition. *Perception* **2002**, *31*, 995–1003. [CrossRef]
118. Bindemann, M.; Burton, A.M. The role of color in human face detection. *Cogn. Sci.* **2009**, *33*, 1144–1156. [CrossRef] [PubMed]
119. Hunt, R.W.G. *The Reproduction of Colour*, 5th ed.; Fountain Press: London, UK, 1995.
120. MacDonald, L.W. Fatal Flaws: Uncertainty in the Interpretation of Colour in CCTV Images. *Ann. Br. Mach. Vis. Assoc.* **2007**, *7*, 1–11.
121. Bachmann, T. Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *Eur. J. Cogn. Psychol.* **1991**, *3*, 87–103. [CrossRef]
122. Bruce, V.; Henderson, Z.; Newman, C.; Burton, A.M. Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol. Appl.* **2001**, *7*, 207–218. [CrossRef]
123. Costen, N.P.; Parker, D.M.; Craw, I. Spatial Content and Spatial Quantisation Effects in Face Recognition. *Perception* **1994**, *23*, 129–146. [CrossRef] [PubMed]
124. Costen, N.P.; Parker, D.M.; Craw, I. Effects of high-pass and low-pass spatial filtering on face identification. *Percept. Psychophys.* **1996**, *58*, 602–612. [CrossRef] [PubMed]
125. Viték, S.; Klíma, M.; Krasula, L. Video compression technique impact on efficiency of person identification in CCTV systems. In Proceedings of the Proceedings—International Carnahan Conference on Security Technology, Rome, Italy, 13–16 October 2014.
126. Utochkin, S. The Principles of CCTV Design in VideoCAD Interframe Compression. 2011. Available online: https://cctvcad.com/Files/the_principles_of_cctv_design_in_videocad_part5.pdf (accessed on 30 October 2021).
127. Stephan, C.N. Estimating the Skull-to-Camera Distance from Facial Photographs for Craniofacial Superimposition. *J. Forensic Sci.* **2017**, *62*, 850–860. [CrossRef]
128. Klima, M.; Fliegel, K. Image compression techniques in the field of security technology: Examples and discussion. In Proceedings of the 38th Annual 2004 International Carnahan Conference on Security Technology, Albuquerque, NM, USA, 11–14 October 2004; IEEE: Albuquerque, NM, USA, 2014; pp. 278–284.
129. Qi, X.; Liu, C. Mitigate compression artifacts for face in video recognition. In Proceedings of the Disruptive Technologies in Information Sciences IV, online. 27 April–8 May 2020; Blowers, M., Hall, R.D., Dasari, V.R., Eds.; SPIE: Bellingham, WA, USA, 2020; p. 25.
130. Ghazali, N.N.A.N.; Zamani, N.A.; Abdullah, S.N.H.S.; Jameson, J. Super resolution combiantion methods for CCTV forensic interpretation. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications, Kochi, India, 27–29 November 2012; pp. 853–858.
131. Stevenage, S.V.; Pitfield, C. Fact or friction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis. *Forensic Sci. Int.* **2016**, *267*, 145–156. [CrossRef]
132. Norell, K.; Läthén, K.B.; Bergström, P.; Rice, A.; Natu, V.; O'Toole, A. The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons. *J. Forensic Sci.* **2015**, *60*, 331–340. [CrossRef] [PubMed]
133. FISWG. *Guide for Role-Based Training in Facial Comparison*; FISWG: Washington, DC, USA, 2020; Available online: https://www.fiswg.org/fiswg_guide_for_role-based_training_in_facial_comparison_v1.0_20200717.pdf (accessed on 30 October 2021).
134. Towler, A.; Kemp, R.I.; Mike Burton, A.; Dunn, J.D.; Wayne, T.; Moreton, R.; White, D. Do professional facial image comparison training courses work? *PLoS ONE* **2019**, *14*, e0211037. [CrossRef] [PubMed]

135. Davis, J.P.; Tamonyte, D. Masters of disguise: Super-recognisers' superior memory for concealed unfamiliar faces. In Proceedings of the 2017 Seventh International Conference on Emerging Security Technologies (EST), Canterbury, UK, 6–8 September 2017; IEEE: Canterbury, UK, 2017; pp. 44–49.
136. Russell, R.; Duchaine, B.; Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **2009**, *16*, 252–257. [CrossRef] [PubMed]

biology

MDPI

Article

Forensic Anthropology as a Discipline

Nicholas V. Passalacqua [1,*], Marin A. Pilloud [2] and Derek Congram [3]

[1] Department of Anthropology and Sociology, Western Carolina University, Cullowhee, NC 28723, USA
[2] Department of Anthropology, University of Nevada, Reno, NV 89557, USA; mpilloud@unr.edu
[3] Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; d_congram@sfu.ca
* Correspondence: nvpassalacqua@wcu.edu

Simple Summary: Forensic anthropology in the United States is a specialization within the overall field of anthropology. Forensic anthropologists are specially educated and trained to search, recover, and examine human remains within a medicolegal context. Over time, forensic anthropology has become increasingly specialized and distinct from other specializations within anthropology. As such, we argue that forensic anthropology should be considered its own discipline, with a unique knowledge base, separate from other similar forms of anthropology, such a bioarchaeology. We argue that forensic anthropologists have unique expertise, making them the only type of anthropologist qualified to perform medicolegal examinations of human remains. Finally, we contend that to perform or represent yourself as a forensic anthropologist without the appropriate expertise is ethical misconduct. The value of this paper is that it explains the importance of expertise and knowledge, and how forensic anthropology has diverged from other specializations of anthropology enough to be considered its own discipline.

Abstract: This paper explores the current state of forensic anthropology in the United States as a distinct discipline. Forensic anthropology has become increasingly specialized and the need for strengthened professionalization is becoming paramount. This includes a need for clearly defined qualifications, training, standards of practice, certification processes, and ethical guidelines. Within this discussion, the concept of *expertise* is explored in relation to professionalization and practice, as both bioarchaeology and forensic anthropology have different areas of specialist knowledge, and therefore unique *expertise*. As working outside one's area of expertise is an ethical violation, it is important for professional organizations to outline requisite qualifications, develop standards and best practice guidelines, and enforce robust preventive ethical codes in order to serve both their professional members and relevant stakeholders.

Keywords: forensic anthropology; bioarchaeology; qualifications; expertise; knowledge; ethics; education; professionalization; standards

Citation: Passalacqua, N.V.; Pilloud, M.A.; Congram, D. Forensic Anthropology as a Discipline. *Biology* **2021**, *10*, 691. https://doi.org/10.3390/biology10080691

Academic Editors: Maria Giovanna Belcastro and Marco Milella

Received: 15 June 2021
Accepted: 15 July 2021
Published: 21 July 2021

1. Introduction

Bioarchaeology and forensic anthropology are two closely related specializations of biological anthropology that examine human remains to understand the life experience and biological parameters of the individuals from which the remains are derived. In this treatment, we focus specifically on these two disciplines as they are closely linked by their study of anatomically modern human skeletal remains. We also limit our focus to the United States, recognizing that there are distinct education, practice, and professional qualification standards in different countries; in part stemming from different national/regional education systems and legal statutes. While forensic anthropology and bioarchaeology have different goals, both disciplines use similar approaches and sometimes the same methods to examine human remains, typically, gross skeletal material (to include bones and teeth) to determine such parameters as species (to ensure the remains are human in origin), sex (sometimes gender in conjunction with other contextual information), age

(sometimes stage-of-life/life history), and stature (living height). Both disciplines also perform more complex analyses examining such characteristics as population variation in terms of biological distance (sometimes called ancestry/population affinity/boaffinity), antemortem and perimortem trauma, taphonomic modifications (sometimes postmortem interval), anomalous and pathological conditions, skeletal indicators of biological stress, and inferential data using archaeological context (sometimes mortuary patterns).

Bioarchaeology and forensic anthropology derive from biological anthropology, but are, at least in their ideal forms as practiced in the model of the United States, hybrids of both biological anthropology *and* archaeology. Both disciplines require the understanding of human bone biology *as well as* archaeological context and taphonomic changes to generate comprehensive conclusions about the lives (and in some cases the death or death event) of individuals. They draw from agency theory regarding the introduction of human remains into the archaeological record usually via culturally intentional actions for various purposes that can reflect culture and cultural identity more broadly [1]. Additionally, both are firmly entrenched in anthropology through their biocultural approach to understanding human biological adaptation, or the interpretation of skeletal modifications (during life, at death, and after death) through a cultural lens.

In the current paradigm, it is not uncommon for an individual trained in one subdiscipline of biological anthropology to offer expertise and services in another, and this is particularly common between bioarchaeology and forensic anthropology. In fact, Ubelaker [2] p. 137, claimed, "[t]he symbiotic and dynamic relationship of these academic areas greatly improves the quality of the applications of each". Contrarily, Juarez [3] argued that a focus on the commonalities between bioarchaeology and forensic anthropology is problematic as it does not emphasize the differences and boundaries of each discipline. Thompson [4] p. 68, agrees that viewing the work of a forensic anthropologist as being easily done by any trained osteologist is "a misperception of what the subject involves through focusing on methods while ignoring context". While, Ross [5] argued that forensic anthropologists are inherently more stringent in analyses and could do all that a bioarchaeologist can do, but not vice versa. While we agree with Ubelaker [2] in that both disciplines benefit from each other, we also agree in *concept* with Juarez, Thompson, and Ross in that both disciplines are becoming increasingly complex and specialized, such that education and training in one discipline do not translate into competency or expertise in the other.

This disagreement has precipitated the need for increased professionalization in terms of standardizing education, defining qualifications, defining and implementing ethical codes, and reconsidering the roles played by professional organizations within both bioarchaeology and forensic anthropology. We view all these issues as interwoven and each influencing the other; however, thus far, they have not been explicitly addressed comprehensively in the literature.

As both bioarchaeology and forensic anthropology have grown significantly in the last few decades, it has become prudent to explore their differences and similarities and the need for their individualization and professionalization in terms of defining qualifications (i.e., education and training needed to demonstrate adequate knowledge to perform discipline-related tasks in an applied setting) and expertise. These topics must also be framed as issues that would best be addressed by professional organizations, as disciplinary leaders harnessing the power of their communities of practitioners. This exercise is not a means of *academic gatekeeping*; but rather a means to identify minimum standards and best practices of what to expect *at a minimum* from an individual practicing a particular profession [6,7].

The goal of this paper is to consider both bioarchaeology and forensic anthropology as unique disciplines, having diverged due to increasing specialization and scholarly distancing; thus, bioarchaeologists and forensic anthropologists have their own unique areas of expertise and spheres of practice. While bioarchaeology and forensic anthropology can vary greatly in their education and practice globally, we focus on the practice of these subdisciplines within the United States. For overviews of forensic anthropology in other countries, there are several excellent treatments to which the reader can refer, e.g., [8–28]. In

this treatment, we begin with a brief discussion on the trend of increasing specialization and decreasing overlap in educational programs and scholarship between the two disciplines. We follow with a definition and discussion on the scholarship of expertise and its relevance to considering bioarchaeology and forensic anthropology as unique expertise. We then expand this discussion within the context of professional qualifications, primarily in regard to the role of ethical codes and professional organizations. Next, we provide an overview of qualifications and their importance in relation to expertise, education, and practice, followed by a discussion of how to codify expertise and practice using best practice recommendations and standards documents, which are becoming ever more popular within the forensic sciences and may soon be required for practice within that context. Finally, we conclude with recommendations for the future, a call for greater consideration of the importance of qualifications as a means of respecting both the remains of those we study and their extant next-of-kin/communities, and provide a glossary (Glossary) defining several of the terms used throughout this paper for standardization and clarification.

The Divergence of Bioarchaeology and Forensic Anthropology

Early versions of both bioarchaeology and forensic anthropology were originally practiced by physicians, anatomists, and biological anthropologists with interests in the examination of the human skeleton. The examination of skeletal remains, as well as the types of research questions addressed, have always been dependent upon the contexts from which the remains were derived. When skeletons within archaeological contexts are excavated, researchers want to know about the life experiences of these individuals. Hypotheses may be formulated to pursue research around migration, diet, stress, violence, social structure, disease loads, activity levels, disability, mortuary practices, fertility, demography, growth and development, and life history, among many others, e.g., [29–41].

When modern skeletons are discovered, anthropologists and the medicolegal community want to know the identity and circumstances of the death of that individual. To pursue identification, they may estimate the individual's biological profile (i.e., age, ancestry/population affinity, sex, and stature), describe individualizing features, and compare ante- and postmortem data, e.g., [42–53]. They are also interested in the circumstances surrounding the death event, illustrated by perimortem trauma and taphonomic alterations, and potentially estimating a postmortem interval, e.g., [54–57]. Research also exists on the applicability of indicators of biological stress as part of the identification process [58] and investigations into gross human rights violations and structural violence [59–61]. However, the collection or analysis of such data is not routinely performed as part of forensic anthropological casework (i.e., reports provided in a medicolegal context).

Over the past several decades, various methods have been developed to best address the research agendas of each discipline, with differing foci based on the context-dependent nature of these investigations. Academically, these differing research agendas have increasingly diverged into academic programs and graduate advisors specializing in bioarchaeological *or* forensic anthropological approaches. In doing so, bioarchaeology and forensic anthropology have slowly deviated in terms of professional conferences attended [62], academic advisors and institutions, bodies of literature, venues of publication, and professional memberships. Through this divergence, they have become more and more isolated from each other, developing separate communities of practice with separate "social lives" [63]. All scientific disciplines essentially function in this manner, effectively "mold[ing] their disciplines by pedagogically fashioning their disciples" [64] p. 381. The choices made by academic hiring committees for future directions of a program are the same as those made by graduate student advisors in that they are purposeful, active choices, which intentionally shape future generations of pedagogy [64–68].

It is unclear precisely when individuals in the United States studying human skeletons from archaeological contexts, being educated in biological anthropology, began to identify as *bioarchaeologists*, or when individuals studying human skeletons in medicolegal contexts being educated in biological anthropology began to identify as *forensic anthropologists*; that

is, as opposed to identifying as biological/physical anthropologists. It is likely that as each subdiscipline increased in popularity, practitioners began to self-identify as one or the other. According to Snow [69] and Tersigni-Tarrant and Langley [70], individuals began identifying as forensic anthropologists in the 1970s; however, this trend grew in the 1980s when forensic anthropology began to gain popularity from the work of William Bass, Walter Birkby, William Maples, and their graduate students (John Williams, personal communication 2019).

In the 1990s, bioarchaeology began to increase in visibility with the passage of the Native American Graves Protection and Repatriation Act (NAGPRA), which required trained osteologists to assist in repatriation efforts. Additionally in the 1990s, two formative publications in bioarchaeology were released: *Standards for Data Collection from Human Skeletal Remains* [71] and Clark Spencer Larsen's [72] *Bioarchaeology: Interpreting Behavior from the Human Skeleton*. It is likely, then, that around the 1980s and 1990s individuals began to identify as either bioarchaeologists or forensic anthropologists, pursuing graduate advisors based on such foci and graduate programs with discipline-specific education programs. While there are some individuals who practice both and who consider themselves a bioarchaeologist and a forensic anthropologist, this number has likely decreased over the past several decades based on our observation of the subdisciplines, and may continue to do so as each becomes more specialized. An increase in full-time applied positions for forensic anthropologists [73] has also surely influenced this trend.

Considering the divergence of bioarchaeology and forensic anthropology, Buikstra et al. [74] found large increases in publications focusing on bioarchaeology *and* forensic anthropology starting in the 1970s, corresponding with the incipient formalization of both areas of study. However, Buikstra et al. [74] also demonstrated that while bioarchaeological literature was found in a variety of anthropological journals, forensic anthropological literature was increasingly published in the *Journal of Forensic Sciences* to the exclusion of other more anthropologically focused journals. Bethard [75] also demonstrated this trend by practicing forensic anthropologists certified by the American Board of Forensic Anthropology (ABFA). Bethard [75] found that based solely on the focus of dissertation subjects, representing the focus of graduate research projects, forensic anthropologists have increasingly pursued forensic anthropological research topics, rather than bioarchaeological or other more general biological anthropology topics, particularly since 2005.

Only recently have the first journals dedicated to bioarchaeology or forensic anthropology been established. Arguably, the first journal dedicated to bioarchaeology was the *International Journal of Osteoarchaeology*, established in 1991 and published by Wiley. Although, like the term osteoarchaeology itself, this journal has a heavy European focus and includes many publications on the analysis of non-human remains. It was not until 2017 that the journal *Bioarchaeology International* was established, published by the University of Florida Press [76]; this journal was followed one year later in 2018 by the first journal dedicated to forensic anthropology, *Forensic Anthropology* [77], also published by University of Florida Press (Gainesville, FL, USA).

Martin [78] p. 163, points out that bioarchaeologists have long been critical of forensic anthropological work as being "merely technical expertise". She fully admits that she used to be one of those bioarchaeologists who questioned "where is the anthropology in forensic anthropology?" [78] p. 163; yet, she has changed her viewpoint on the topic in a recognition that forensic anthropology is not an atheoretical practice. Martin is more hopeful on the integration of the interests of bioarchaeology and forensic anthropology, and has been involved in editing volumes that promote this integration of bioarchaeological and forensic anthropological approaches to research questions, e.g., [79,80]. She offers the term "forensic bioarchaeologist" as a means to promote this cross-disciplinary effort [78]. The term "forensic bioarchaeologist" was previously introduced by Scott and Connor [81], Skinner and colleagues [82], and Jessee and Skinner [83]. The latter two used it as a means of integration of archaeological methods and theory into medicolegal investigations of mass graves. Of particular pertinence to this discussion, Skinner et al. (2003) promoted

guidelines for bioarchaeological practice in a forensic context. Nevertheless, the term "forensic bioarchaeologist", has generally not been adopted.

Conversely, Steadman [84] discourages the use of such a term, arguing that it may serve to obscure the lines between forensic anthropology and bioarchaeology. In an academic sense, this may be "harmless" [84] p. 251; however, the term may cause confusion in the public about the distinction between the two disciplines, which she feels could potentially be problematic for jurors. We too argue the term could blur the boundaries of qualifications and expertise between the two disciplines, which is challenging for law enforcement and those working in the medicolegal realm. The further confounding of the two subdisciplines is also evidenced by the *American Journal of Physical Anthropology's* manuscript submission system. In this system, authors must designate a manuscript by "subfield", with the only relevant choice for bioarchaeology or forensic anthropology being: "Bioarchaeology [including forensics]".

We agree with Martin [78] p. 163, and Ubelaker [2] (as discussed previously) that the subdisciplines of bioarchaeology and forensic anthropology are independent and *complementary* and while they may differ in focus, contextual application, and specific hypotheses, they can learn much from each other. We firmly believe that a clear standardization of education, training, and qualifications is the best way for this mutual appreciation and professionalization to be achieved. The first step in this process is recognizing a lack of cross-disciplinary expertise, which can be achieved through a broader understanding of what constitutes expertise, as we discuss further below.

2. Disciplinary Expertise

As we argued above and as others have demonstrated [74,85], while lacking published qualifications, bioarchaeology and forensic anthropology have developed into their own disciplines each with their own areas of expertise, bodies of literature, analytical methods, theoretical models, and education programs. However, it is important to discuss what expertise is and how it is created to fully appreciate the implications of differing expertise (and thus different disciplinary skills). Typically, we consider experts to have *authoritative* knowledge or skill in a particular area, while laymen are non-professional individuals lacking expertise [29,30,86–88]. A depth of literature has emerged relatively recently examining experience, expertise, and the sociology of scientific knowledge, e.g., [63,86,88–107]. We include a brief discussion of this literature here for some of the same reasons Collins [93] p. 127, was motivated to pioneer this avenue of inquiry, "to persuade sociologists [here, anthropologists] to reflect upon their expertises".

Collins and Evans [108], present models of various forms of specialist expertise along a two- or three-dimensional spectrum [97,98,100]. Within specialist expertise exist two main types of knowledge, "Ubiquitous Tacit Knowledge" and "Specialist Tacit Knowledge" (Table 1). The first, "Ubiquitous Tacit Knowledge" (i.e., information) may be generated simply via reading without interacting with the appropriate contributory experts, this is knowledge that is easily accessible and therefore common knowledge. The novice level of "Ubiquitous Tacit Knowledge" is considered "beer mat" (knowledge of very superficial facts about a topic that you might find on a beer mat/coaster). The next level is "popular understanding", which can be achieved through popular non-fiction books and general media. "Primary source knowledge" involves engaging with the primary literature; however, this literature still only provides "a shallow or misleading appreciation of science in deeply disputed areas", which is far from obvious for the uninitiated [108] p. 22.

"Specialist Tacit Knowledge" must be acquired via interactions and enculturation with practicing professionals [100], and serves as the necessary knowledge base(s) to practice a discipline. Specialist Tacit Knowledge ranges from "interactional expertise" to "contributory expertise" [108]. Interactional expertise is essentially the ability to interact with other experts using their language/jargon and understanding the concepts being discussed, but lacking the full expertise to practice [102]. Contributory expertise is traditional technical expertise, where the practitioner is the contributory and interactional expert, meaning they

are able to discuss/interact with other individuals at a complex level *and* able to perform complex disciplinary tasks competently [105]. With these definitions of types of knowledge, expertise can be defined as "the mastery of the tacit knowledge of a domain of practice, with interactional expertise being mastery of the domain's language, and contributory expertise being the ability to competently engage in the practices of that domain" [104] p. 99.

Table 1. Levels of expertise based on Collins and Evans (2007).

	UBIQUITOUS TACIT KNOWLEDGE			**SPECIALIST TACIT KNOWLEDGE**	
SPECIALIST EXPERTISES	**Knowledge That Is Easily Accessible (i.e., Ubiquitous)**			**Exclusive Knowledge That Must Be Acquired via Interactions and Enculturation with Practicing Contributory Experts**	
	Beer Mat Knowledge	**Popular Understanding**	**Primary Source Knowledge**	**Interactional Expertise**	**Contributory Expertise**
	Knowledge of very superficial facts about a topic such that you might find on a beer mat/coaster	Knowledge based on popular non-fiction books and the general media	Knowledge based on engaging with the primary literature.	This represents having enough expertise about a discipline to interact with its contributory experts performing their work, but lacking the technical knowledge to perform it yourself.	This represents having enough expertise to contribute to a discipline through its technical and scholarly practice
			Note that literature still only provides "a shallow or misleading appreciation of science in deeply disputed areas" (Collins and Evans 2007:22)	"Scientists themselves tend to have contributory expertise in their narrow specialism and interactional expertise in cognate specialisms". (Collins 2004:141)	

As both bioarchaeology and forensic anthropology share many common lower-level knowledge areas, individuals educated in either discipline have some specialist knowledge of the other, representing what Collins and Evans [108] refer to as Primary Source Knowledge. For example, both may use the same method to estimate the sex of skeletal remains. However, as specialization increases, there is less and less overlap in knowledge, and the expertise required to interpret method results and generate reports becomes more exclusive. For example, bioarchaeologists must understand the historical context of the samples they are analyzing and when possible, work with descendent communities; while forensic anthropologists must understand jurisdiction, chain of custody, and admissibility of evidence.

These distinctions in knowledge area and specialist expertise are important, as without the appropriate "Specialist Tacit Knowledge", practitioners/researchers may perform tasks inappropriately and/or incorrectly. As Collins and Evans [108] p. 22, state: "it can be shown that what is found in the literature, if read by someone with no contact with the core-groups of scientists who actually carry out the research in disputed areas, can give a false impression of the content of the science as well as the level of certainty". In other contexts, this concept is often referred to as the Dunning–Kruger effect, or essentially the ignorance of one's ignorance [109–112]. Individuals have just enough knowledge to understand the primary literature, but not enough to fully grasp the nuances of this material or how to properly discuss or apply it. The implications of which are that bioarchaeologists and forensic anthropologists, as contributory experts in their respective disciplines, can be ignorant of their lack of cross-disciplinary expertise. Collins and Evans state: "Enculturation" is the only way to master an expertise which is deeply laden with tacit knowledge because it is only through common practice with others that the rules that cannot be written down can come to be understood" [108] p. 24. Essentially, as knowledge becomes more specialized, individuals interested in acquiring this knowledge must rely on practitioners' practices (i.e., experiential training programs) rather than literature (i.e., educational pro-

grams) [104]. Returning to bioarchaeology and forensic anthropology, essentially, the only way to develop contributory expertise in one of these disciplines is through enculturation in a bioarchaeological or forensic anthropological educational and/or training program supervised by a contributory expert in that discipline. This is not to say that individuals cannot be experts in both disciplines, rather it means that dual expertise requires individuals to develop contributory expertise in *both* bioarchaeology and forensic anthropology. As Collins and Evans [108] point out, lacking such enculturation at the level of contributory expertise leads to overconfidence and poor performance (i.e., the Dunning–Kruger effect).

It is important to reiterate here that the focus is on education and training by other contributory experts, working towards building a body of knowledge and practical skills in a way that is consistent with how the discipline (i.e., other contributory experts) operates. This "enculturation" is not a form of limiting access to knowledge, but rather as means of acquiring knowledge in such a way that the learner will develop interactional expertise (being able to have high-level discussions with other contributory experts, using the appropriate processes) and contributory expertise (being able to use methods, technology, etc., in such a way that it contributes to the greater body of scholarly knowledge of a discipline). This is not a new concept and is essentially how academia currently operates. That is, students attend graduate school at programs that have education programs in which they are interested, working with advisors doing work similar to what they want to do as professionals. While academia is not without its major flaws, the argument here is simply that training and education are critical to gaining the requisite skills to perform disciplinary tasks. The arguments presented here are the first step in recognizing the need for developing expertise, the next step would be to develop such training and education programs. As a discipline, we can and should critically evaluate what this training looks like and how we define demonstrable expertise in a way that is inclusionary and equitable.

3. Ethics, Expertise, and the Role of Professional Organizations

Professional organizations like the American Academy of Forensic Sciences (AAFS), the American Association of Physical/Biological Anthropologists (AAPA/AABA), the Society for American Archaeology (SAA), and the American Anthropological Association (AAA) exist largely to provide individuals in that profession opportunities to network, organize, and serve and engage with the public. Additionally, these organizations typically provide professional development and continuing education opportunities, which is why student members are often encouraged to join as a means to facilitate disciplinary enculturation and entrée into the profession at large, which can also serve to provide them with the necessary expertise to practice their discipline. Further, organizational ethical codes should address qualifications so as to define the proper education and/or training to perform discipline-related tasks. Such a definition would prevent an individual from performing applied work outside their area of expertise, which is an ethical violation. Here, we first outline the need for professional ethical guidelines, and then we revisit the role these organizations can play in providing qualifications for members.

4. Why Do We Need Professional Ethical Guidelines?

A professional is someone who: (1) possesses a body of special knowledge (i.e., contributory expertise), (2) practices within an ethical framework (i.e., adheres to a code of ethics and avoids conflicts of interest), and (3) fulfills a societal need [6,7,113]. Professionalism is conduct associated with a particular profession. Both bioarchaeology, e.g., [114–119] and forensic anthropology, e.g., [120–124] have extensive literature regarding ethical conduct and practice. However, ethical codes are typically established, and presumably enforced, by professional organizations. Ethical codes are used to: (1) establish conduct that is meant to be pursued by practitioners of a discipline (altruistic behavior); (2) establish conduct that must be avoided (i.e., misconduct), and (3) provide potential negative outcomes for practitioners engaging in misconduct [113,125]. Professional ethics are typically presented in the form of either aspirational guidelines or preventive standards [126]. Aspirational ethical

codes are meant to promote human wellbeing and present a number of guiding and/or motivational behaviors that an organization *would like* its members to follow/achieve. Preventive ethical codes are enforced by an adjudicating committee within an organization that performs an investigation when a complaint of misconduct alleges that an individual acted unethically [113]. The content of ethical codes for professional organizations vary, but should generally be structured to ensure members avoid unprofessional conduct, so as to maintain the credibility of the profession and professional organization. Without clear professional ethics, a discipline does not have guidelines for acceptable or unacceptable behavior, such that there can be no good or bad conduct, and all actions must be treated equally [127] p. 233.

In terms of the meaningful implementation of professional ethics, there are two essential issues that must be addressed. The first is that ethical codes must be detailed enough so that specific types of conduct considered to be unethical are demonstrably so. Second, ethical codes must be enforceable, with negative outcomes for individuals found guilty of misconduct. The importance of these issues is perhaps most easily demonstrable in terms of U.S. politics, where ethical guidelines are often ignored when ethical misconduct is not actually against the law, and the language of ethical guidelines is vague and not rigorously enforced [128–130]. As Josephson states: "there is a big difference between what you have the right to do and what is right to do" [131] p. 4. Unfortunately, the same is also the case in most professional organizations in which bioarchaeologists and forensic anthropologists are members. This is important as the law is meant to represent and enforce values for society as a whole, but is often not specific enough to cover many activities pertinent to a particular profession. Professional ethical codes more directly address discipline-specific values and behaviors.

Because ethical codes are tied to specific organizations, they only apply to the members of those organizations. This means that organizational membership (or lack thereof) plays an important role in establishing and policing the conduct of a profession and its body of practitioners, based on each organization's ethical code. It also means that each organization should consider the ramifications of its membership requirements in terms of professional qualifications and access to students and non-experts, and how this allows the organization to serve its role within its professional community. Therefore, professional organizations serve a role of providing opportunities for gaining expertise through *enculturation* by interacting with additional contributory experts, and are positioned to provide sanctions when a member practices outside of their expertise, which could be seen as an ethical violation.

5. The Need for Disciplinary Qualifications

Bioarchaeology has no official or widely accepted published documents in the public or private sector regarding qualifications for bioarchaeological practice, or for the education or training of its practitioners. Currently, anyone *claiming* to have the appropriate training in bioarchaeology can be employed to perform applied bioarchaeological tasks. This is particularly true in contract archaeology (i.e., cultural resource management [CRM]), where it may be difficult to find qualified osteologists who are also available at the time of the excavation. These companies may then be forced to hire individuals with little osteological training to excavate and perform analyses.

As a recognition of the need for standardized qualifications, there have been some movements to define minimum qualifications to perform osteological analysis and excavation. Within the Code of Colorado Regulations, under Section 13 "Unmarked Human Graves", point G states, "Pursuant to 24-80-1302(4)(e), the physical anthropological study of human remains shall be conducted by a qualified physical anthropologist with the credentials comparable to those required for principal investigators, as set forth in Section 5 of these regulations" (https://www.sos.state.co.us/CCR/GenerateRulePdf.do?ruleVersionId=541&fileName=8%20CCR%201504-7, accessed on 16 July 2021). The qualifications outlined for principal investigators include a graduate degree in anthropology,

archaeology, or history with experience in Colorado archaeology; one year of professional experience; four months of supervised field and analytic experience; and the ability to complete research.

The Wisconsin Historical Society has taken this a step further to establish specific guidelines to be a "qualified archaeologist for burial sites or a qualified skeletal analyst". Their mandatory requirements include a graduate degree in anthropology, at least one year of professional experience or specialized training, at least four months of supervised experience, and the ability to complete a project. The full list of requirements and application instructions can be found on their website: https://www.wisconsinhistory.org/Records/Article/CS14963, accessed on 16 July 2021. The Society for California Archaeology (SCA) recently sent out an email to members with a draft outline for "recommended qualifications for field osteologists working in California". Very generally, this guideline would recommend a master's degree in anthropology and field experience dealing with human remains. There are additional qualifiers such as course work in human osteology, experience with NAGPRA, and a field school with an osteological focus, among others. The guidelines have not been published and are currently out for public comment (https://form.jotform.com/90855960158972, accessed on 16 July 2021). Of note, the guidelines would not be enforceable by the SCA, but would serve as recommendations for employers. These guidelines for qualifications address specialist expertise by requiring not only advanced education but also having already worked as a professional and having been supervised to gain enculturation.

Within forensic anthropology, Galloway and Simmons [132] identified deficiencies in the standardization of education and training in forensic anthropology over two decades ago. As a result, more formal efforts were taken up by the Scientific Working Group for Anthropology (SWGAnth) to establish guidelines for Qualifications [133], and Education and Training [134]. However, these documents were never widely adopted, nor are they enforceable. Passalacqua and Pilloud [85] surveyed practicing forensic anthropologists and found large variations in terms of graduate coursework taken by the survey participants. The survey also demonstrated a lack of consensus among practicing forensic anthropologists in what constituted appropriate education and training in forensic anthropology. However, there was overwhelming agreement that clear standards for education are needed, with a high degree of support (96%) for developing an accreditation for forensic anthropology educational programs. Additionally, Langley and Tersigni-Tarrant [135] outlined a model to develop qualifications in forensic anthropology based on medical education. In this model, there would be a set of core competencies demonstrated via various "entrustable professional activities". Once core competencies are clearly identified and agreed upon, appropriate training and certification (to demonstrate expertise) could be implemented.

There are currently no certifications for the profession of bioarchaeology, however, job ads in the United Kingdom for osteologists have added "professional grade membership of the CIFA [Chartered Institute for Archaeologists]" as a desired criterion; which is functionally a certification (albeit not necessarily focused on the analysis of human remains).

While no widely supported guidelines or standards currently exist in terms of education and training, or qualifications within forensic anthropology, there are certification processes see, [7] for an overview. In Europe, the Forensic Anthropology Society of Europe (FASE) and the Royal Anthropological Institute (RAI) oversee certification processes. In Latin America, the Asociación Latinoamericana de Antroplogía Forense (ALAF) also has a certification process. In the United States, the ABFA has a certification process, and is currently the only accredited certifying body for forensic anthropology.

While these certifications may exist, there is still a lack of clear qualifications (i.e., who can practice and how do you educate/train practitioners) within *both* subdisciplines. This lack of standardized qualifications is problematic as there are no widely agreed-upon standards for determining who is and who is not an expert and thus capable to perform tasks as a bioarchaeologist or a forensic anthropologist. As both disciplines are specialized, it can be difficult for outside agencies to accurately judge the qualifications and requi-

site expertise of individuals applying to perform these types of analysis. For example, with very few exceptions there are no standards in bioarchaeology to determine who can perform work for NAGPRA repatriation, osteological analysis, or excavation in CRM, or at academic archaeological sites. Nor are there requirements demonstrating expertise to teach bioarchaeological field-schools, or meaningful certifications or competencies gained through attending a field school. All of these things can be problematic for the adequate interpretation of archaeological sites and human remains with an irreplaceable loss of data and information (when analyses are permitted), particularly in cases of repatriation and reburial. Moreover, while there is a certification process within forensic anthropology, there is no legal requirement that a forensic anthropologist must be certified in order to perform such analyses in a medicolegal context. In fact, any such self-identified specialist can be tasked to perform this work. Contracting unqualified individuals to perform forensic anthropological casework can result in improper conclusions, which can hinder identification efforts (or worse, misidentify a person) and could have enormous consequences during the litigation process, for the analyst, their employer and, critically, for the family of the deceased.

The ramifications of the differences in the disciplines are that if an individual without the appropriate education and training acts beyond their professional expertise, they are misrepresenting their qualifications and could potentially do harm to the research project, field recovery, forensic case/investigation, descendant populations (as occurred with the well-known example of "The Ancient One", also known as Kennewick Man [136–138], and/or the entire discipline; and are thus acting unethically. As such, professionals and professional organizations should be critically concerned about qualifications, expertise, and ethical practice. It cannot be expected that law enforcement agencies, attorneys, CRM firms, museums, or a medicolegal authority be trained in examining the nuances of an anthropological degree to determine who is and who is not qualified to be a bioarchaeologist or forensic anthropologist. There must be clear published standards that go beyond education in skeletal analysis as work in bioarchaeology and forensic anthropology has become increasingly specialized and individuated.

6. Conclusions: A Way Forward

We attempted to demonstrate that bioarchaeology and forensic anthropology have evolved and diverged into two separate disciplines, each encompassing its own suite of literature, education, training, and qualifications. Additionally, we attempted to illustrate the connection between expertise, ethics, and professional organizations as important aspects to the advancement of, and professional practice in, both bioarchaeology and forensic anthropology. Both disciplines are in need of the development of standardized education and training programs that reinforce best practice models for their applied foci. Once appropriate models for education and training have been defined, it can become possible for practitioners to demonstrate expertise to achieve credentials in a more meaningful and demonstrable way. Professional organizations should be leading these efforts in addition to establishing robust and enforceable ethical codes and tailoring their membership in such a way as to support the further professionalization of their disciplines.

Thus far, bioarchaeology and forensic anthropology have not fully embraced standardization of practice or qualifications—although forensic anthropology is ahead in this regard [7]. This shortfall is probably due in part to the largely academic focus of both bioarchaeology and forensic anthropology and the lack of familiarity with program accreditation in anthropology generally (although this is common, if not required, in many other academic disciplines, including many of the sciences). However, the accreditation of academic programs specializing in bioarchaeology or forensic anthropology may be a relatively straightforward way to ensure the generation of expertise and qualifications. The generation of consensus-based qualifications (via graduate-level education) for these subdisciplines would not necessitate large changes in curricula within anthropology departments. Rather, these programs may need to make small adjustments to fit the required

definitions for accreditation. For example, the definition of qualifications to be a bioarchaeologist could be graduate courses in osteology, bioarchaeology, archaeological theory, and a field school; courses that many bioarchaeologists would readily take and are currently offered by graduate programs. Further, this coursework could be modeled to allow for the development of competency of various related applied skills. For example, the osteology course could provide modules on human vs. non-human identification, or the field school could provide verification of expertise to adequately excavate and document skeletal material within an archaeological context. For forensic anthropology, graduate coursework with a forensic focus, osteology, and a forensic archaeological field school could be required. Competency could be demonstrated via mock or mentored casework.

There are already movements to define qualifications within bioarchaeology and forensic anthropology; however, these should be codified by professional organizations and linked to education and training. Within bioarchaeology, there are regional movements to define qualifications to perform bioarchaeological work, but these efforts are in their earliest stages. As there is currently no professional organization for bioarchaeology, these steps are being taken by other organizations, such as the SCA, the Register of Professional Archaeologists (RPA), and the state governments in which the work is being performed. These qualifications are still vague and may not be broadly enforced or accepted by the professional community at large. It may be necessary for this work to be undertaken as a working group within the AAPA/AABA, or independently as is being done with forensic anthropology via the Organization of Scientific Area Committees for Forensic Science (OSAC) and the American Academy of Forensic Science Standards Board (ASB). Again, the definition of these qualifications would not serve as a means to hamper research or scientific integrity or as a means of keeping people out of the disciplines, but would aid in determining who is capable of performing disciplinary tasks in an applied setting. It may also be beneficial for bioarchaeology to develop a national certification process similar to the ABFA, which could serve to illustrate requisite expertise or competencies to employers and stakeholders.

Within forensic anthropology, the OSAC and ASB are developing and have published best practice recommendations and standards for performing various disciplinary tasks. As part of this initiative, there is a consideration for education and training, and qualifications; however, these specific documents are not yet finalized. There is already a mechanism to review and approve education programs within the forensic sciences, the Forensic Science Education Programs Accreditation Commission (FEPAC). However, this organization does not currently oversee forensic anthropology programs. Still, the FEPAC model could provide a way forward for accrediting forensic anthropology (and bioarchaeology) educational programs, if necessary. Additionally, the ABFA certification process could be updated and improved to adequately illustrate competencies as outlined by the OSAC and ASB documents.

While this work is being undertaken to improve and standardize both disciplines, there is still very little consequence for not following existing standards or ethical codes. Platitudes on misconduct without operational enforcement mechanisms are not useful. As performing work beyond one's qualifications is unethical, we argue that professional organizations need to more clearly define ethical codes with enforceable consequences.

Bioarchaeology and forensic anthropology do not operate in a vacuum. Both disciplines examine the remains of deceased humans, and every action, use for, and result from that examination affects the beneficiaries and stakeholders associated with those human remains [139]. While not commonly considered, there has been a recent push to acknowledge that the dead retain their humanity and thus should be granted rights [113,140–143]. Additionally, human subjects have next-of-kin, be they direct living relatives, or more distantly related descendant-communities. As those performing these analyses are often responsible for the curation/custody of these remains, we must acknowledge that we have an ethical and moral responsibility to act in the best interests of these individuals and their next-of-kin (usually defined in open dialogue with relatives and kin). The analyses

we perform are used to reach conclusions that are presented in reports, publications, and other media, available to not only the research subject's next-of-kin, but also the public at large. Bioarchaeological reports may be used to understand past human lifeways and as one part of the process for repatriation and return to descendent communities. Forensic anthropology reports may be used not only to bring closure to a family, but in court to adjudicate innocence or guilt. The conclusions of our work have meaning and reflect upon the identity and lives of the deceased as well as the communities from which they came.

When we consider the importance of this type of work, we should want to ensure that the individual performing an analysis is an expert, and we owe that commitment to our communities and the individuals we study. Incorrect analyses and erroneous conclusions cause harm. As such, we should want to ensure that our students are being educated in the methods and skills necessary to perform their work and best serve not just their discipline, but their research subjects. Further, we should want to ensure that the quality of work we perform as professionals continues to be held to a high standard by generating and following best practices and/or standards for analytical decisions in applied practice. We should do these things, not just because they are the right thing to do, but as a service to those upon which we base not only our work, but also our entire careers.

Author Contributions: Conceptualization, N.V.P. and M.A.P.; writing—original draft preparation, N.V.P., M.A.P., D.C.; writing—review and editing, N.V.P., M.A.P., D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Glossary

Accreditation	A credential used to demonstrate that an organization (e.g., a university, medical examiner's office, forensic anthropology laboratory) meets a set of published standards [7]. Most forensic anthropology laboratories are accredited under ISO/IEC 17020 or 17025.
Best Practice	Procedures, methods, and/or techniques that have been accepted as preferable over others as they produce superior results and comply with legal and/or ethical requirements.
Beneficiaries	agents (e.g., victims, families, communities, NGOs) that can be considered interested parties (i.e., stakeholders) in the investigation, benefiting from the forensic services provided and the resolution of the investigation to varying degrees [139].
Certification	A credential provided by a professional organization demonstrating that an individual has met the knowledge and/or skills required to pass their certification process [7]. For example, the ABFA provides a certification and is accredited by the Forensic Specialties Accreditation Board.
Competency	The application of knowledge, skills, and abilities to correctly complete specific disciplinary *tasks* [7].
Contributory Expertise	Traditional technical expertise, where the practitioner is the contributory and interactional expert [105].
Credential	Verifiable document used to demonstrate completion of education and/or training (e.g., transcripts, licenses). Frequently used to "acknowledge, restrict, or protect the use of a title, and/or activities" [7,144] p. 220.
Education	Formal academic coursework from an accredited school, college, or university, resulting in a degree [113,145].
Expert	An individual possessing authoritative knowledge or skill in a particular area, which can be demonstrated via credentials and/or certification [7,87] pp. 29–30.
Expertise	The mastery of not only the salient, but also the tacit, areas of knowledge of a field of inquiry, which includes the ability to use the language of the field of inquiry as well as to engage fully in its practices [104].

Guideline	(i.e.: best practice documents) Published documents providing *recommendations* for how to perform a particular action or process. Guidelines are typically vetted and published by accredited organizations. Their content must be based on practitioner and stakeholder consensus. Guidelines are typically more detailed/descriptive than standards, but are also open to interpretation [7].
Interactional Expertise	The ability to interact with other experts using their language/jargon and understand the concepts they are discussing [102].
Qualifications	Education and training needed to demonstrate adequate knowledge to perform discipline-related tasks in an applied setting [7].
Should	Something that is not mandatory, but is professionally considered best practice.
Shall	A practice that is professionally considered mandatory.
Specialist Tacit Knowledge	Serves as the requisite knowledge base(s) to practice a discipline [100].
Standard	(i.e.: formal standards) Published documents providing *mandatory* rules for how to perform a particular action or process. Standards are typically vetted and published by accredited organizations. Their content must be based on practitioner and stakeholder consensus [7]. Any deviation from published standards can be considered poor practice and a breach of ethics.
Training	Formal structured process of teaching and assessment at a laboratory or other non-educational institution, often resulting in a certificate [7,113,145].
Ubiquitous Tacit Knowledge	Knowledge that is easily accessible and therefore ubiquitous [100].

References

1. Boyd, C.; Baden, W. Archaeological Inference and Its Application to Forensic Anthropology. In *Forensic Anthropology: Theoretical Frameworks and Scientific Basis*; Boyd, C., Boyd, D., Eds.; Wiley: Hoboken, NJ, USA, 2018; pp. 273–295.
2. Ubelaker, D.H. The Dynamic Interface of Bioarchaeology and Forensic Anthropology. *Intersecc. Antropol.* **2016**, *17*, 5–38.
3. Juarez, C. Commentary 2: Does A Focus On Commonalities Obscure Critical Differences? *Intersecc. Antropol.* **2016**, *17*, 148–149.
4. Thompson, T. Deconstructing The Ideal of Standardization In Forensic Anthropology. In *Disturbing Bodies: Perspectives on Forensic Anthropology*; Crossland, Z., Joyce, R.A., Eds.; School for Advanced Research Press: Santa Fe, NM, USA, 2015; pp. 63–84.
5. Ross, A.H. A Forensic Anthropologist Can Be A Bioarchaeologist, But Should A Bioarchaeologist Be A Forensic Anthropologist? *Intersecc. Antropol.* **2016**, *17*, 152–153.
6. Pellegrino, E.D. Professionalism, Profession and The Virtues of the Good Physician. *Mt. Sinai J. Med.* **2002**, *69*, 378.
7. Passalacqua, N.V.; Pilloud, M.A. The Need to Professionalize Forensic Anthropology. *Eur. J. Anat.* **2021**, *25*, 35–47.
8. Cox, M. Forensic Anthropology and Archaeology: Past and Present—A United Kingdom Perspective. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 29–41.
9. Cattaneo, C. Forensid Anthropology and Archaeology: Perspectives From Italy. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 42–48.
10. Baccino, E. Forensic Anthropology: Perspectives From France. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 49–55.
11. Prieto, J.L. A History of Forensic Anthropology In Spain. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 56–66.
12. Fondebrider, L. The Application of Forensic Anthropology to Invstigation of Cases of Political Violence: Perspectives from South America. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 67–75.
13. Skinner, M.; Bowie, K. Forensic Anthropology: Canadian Content and Contributions. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 87–103.
14. Donlon, D. The Development and Current State of Forensic Anthropology: An Australian Perspective. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 104–114.
15. Indriati, E. Historical Perspectives On Forensic Anthropology In Indonesia. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 115–125.
16. L'Abbe, E.N.; Steyn, M. The Establishment and Advancement of Forensic Anthropology In South Africa. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; John Wiley & Sons, Ltd.: Malden, MA, USA, 2012; pp. 626–638.
17. Marquez-Grant, N.; Litherland, S.; Roberts, J. European Perspectives and The Role of The Forensic Archaeologist In The UK. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; John Wiley & Sons, Ltd.: Malden, MA, USA, 2012; pp. 598–625.
18. Kranioti, E.; Paine, R. Forensic Anthropology In Europe: An Assessment of Current Status and Application. *J. Anthropol. Sci* **2011**, *89*, 71–92. [PubMed]
19. Monteiro Da Silva, S.F.S. Forensic Anthropology and Archaeology In Brazil. In *Biological Anthropology of Latin America: Historical Development and Recent Advances*; Ubelaker, D., Colantonio, S.E., Eds.; Smithsonian Scholarly Press: Washington, DC, USA, 2019; pp. 31–40.

20. Marquez Morgin, L. Forensic Anthropology In Mexico. In *Biological Anthropology of Latin America: Historical Development and Recent Advances*; Ubelaker, D., Colantonio, S.E., Eds.; Smithsonian Scholarly Press: Washington, DC, USA, 2019; pp. 79–88.
21. Monteiro Da Silva, S.F.S. Forensic Anthropology In Northwestern South American (Colombia, Venezuela, Ecuador, and Peru). In *Biological Anthropology of Latin America: Historical Development and Recent Advances*; Sanabria-Medina, C., Osorio Restrepo, H., Eds.; Smithsonian Scholarly Press: Washington, DC, USA, 2019; pp. 238–248.
22. Fondebrider, L. The Development of Forensic Anthropology In Argentina, Chile, and Uruguay: A Brief History. In *Biological Anthropology of Latin America: Historical Development and Recent Advances*; Ubelaker, D., Colantonio, S.E., Eds.; Smithsonian Scholarly Press: Washington, DC, USA, 2019; pp. 325–330.
23. Go, M.C.; Tallman, S.D.; Kim, J. Advances In Forensic Anthropological Research In East and Southeast Asia. *Forensic Anthropol.* **2019**, *2*, 197–203. [CrossRef]
24. Baryah, N.; Krishan, K.; Kanchan, T. The Development and Status of Forensic Anthropology In India: A Review of The Literature and Future Directions. *Med. Sci. Law* **2019**, *59*, 61–69. [CrossRef]
25. Obertová, Z.; Adalian, P.; Baccino, E.; Cunha, E.; De Boer, H.H.; Fracasso, T.; Kranioti, E.; Lefévre, P.; Lynnerup, N.; Petaros, A. The Status of Forensic Anthropology In Europe and South Africa: Results of The 2016 FASE Questionnaire On Forensic Anthropology. *J. Forensic Sci.* **2019**, *64*, 1017–1025. [CrossRef]
26. Bielous, S.D. (Ed.) *Forensic Anthropology Teams In Latin America*; Routledge: London, UK, 2019.
27. Go, M.C. Appraising Forensic Anthropology in The Philippines: Current Status and Future Directions. *Forensic Sci. Int.* **2018**, *288*, E321–E329. [CrossRef]
28. Vaswani, V.; Ahmed, M.N. Forensic Anthropology Education and Training In India. *Med. Sci. Law* **2020**, *60*, 83–84. [CrossRef] [PubMed]
29. Beauchesne, P.; Agarwal, S.C. (Eds.) *Children and Childhood In Bioarchaeology*; University Press of Florida: Gainesville, FL, USA, 2018.
30. Brickley, M.B.; Ives, R. *The Bioarchaeology of Metabolic Bone Disease*; Elsevier: Oxford, UK, 2010.
31. Lewis, M.E. *The Bioarchaeology of Children: Perspectives from Biological and Forensic Anthropology*; Cambridge University Press: Cambridge, UK, 2007; Volume 50.
32. Martin, D.L.; Frayer, D.W. (Eds.) *Troubled Times: Violence and Warfare in the Past*; Routledge: New York, NY, USA, 1997.
33. Allen, M.W.; Jones, T.L. (Eds.) *Violence and Warfare Among Hunter-Gatherers*; Left Coast Press: Walnut Creek, CA, USA, 2014.
34. Agarwal, S.C.; Glencross, B.A. Building A Social Bioarchaeology. In *Social Bioarchaeology*; Agarwal, S.C., Glencross, B.A., Eds.; Wiley-Blackwell: Hoboken, NJ, USA, 2011; pp. 1–11.
35. Geller, P.L. *Bioarchaeology of Socio-Sexual Lives*; Springer: Cham, Switzerland, 2017.
36. Byrnes, J.F.; Muller, J.L. (Eds.) *Bioarchaeology of Impairment and Disability: Theoretical, Ethnohistorical, and Methodological Perspectives*; Springer: Cham, Switzerland, 2017.
37. Tilley, L.; Schrenck, A. *New Developments in The Bioarchaeology of Care*; Springer: Cham, Switzerland, 2017.
38. Tung, T.A. (Ed.) *Violence, Ritual, and The Wari Empire: A Social Bioarchaeology of Imperialism in the Ancient Andes*; University Press of Florida: Gainesville, FL, USA, 2012.
39. Knüsel, C.; Smith, M. (Eds.) *The Routledge Handbook of The Bioarchaeology of Human Conflict*; Routledge: New York, NY, USA, 2014.
40. Pilloud, M.A.; Hefner, J.T. (Eds.) *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*; Academic Press: San Diego, CA, USA, 2016.
41. Martin, D.L.; Harrod, R.P. (Eds.) *The Bioarchaeology of Violence*; Univerity Press of Florida: Gainesville, FL, USA, 2012.
42. Dirkmaat, D.C. (Ed.) *A Companion to Forensic Anthropology*; Wiley-Blackwell: Malden, MA, USA, 2012.
43. Christensen, A.M.; Passalacqua, N.V.; Bartelink, E.J. *Forensic Anthropology: Current Methods and Practice*, 2nd ed.; Elsevier: San Diego, CA, USA, 2019.
44. Tersigni-Tarrant, M.; Shirley, N.R. (Eds.) *Forensic Anthropology: A Comprehensive Introduction*; CRC Press: Boca Raton, FL, USA, 2017.
45. Latham, K.E.; Bartelink, E.J.; Finnegan, M. (Eds.) *New Perspectives in Forensic Human Skeletal Identification*; Academic Press: London, UK, 2018.
46. Blau, S.; Ubelaker, D. *Handbook of Forensic Anthropology and Archaeology*; Routledge: London, UK, 2016.
47. Christensen, A.M. Testing The Reliability of Frontal Sinuses In Positive Identification. *J. Forensic Sci.* **2005**, *50*, 18–22. [CrossRef]
48. Adams, B.J. Establishing Personal Identification Based On Specific Patterns of Missing, Filled, and Unrestored Teeth. *J. Forensic Sci.* **2003**, *48*, 487–496. [CrossRef]
49. Stephan, C.N.; Winburn, A.P.; Christensen, A.F.; Tyrrell, A.J. Skeletal Identification By Radiographic Comparison: Blind Tests of A Morphoscopic Method Using Antemortem Chest Radiographs. *J. Forensic Sci.* **2011**, *56*, 320–332. [CrossRef] [PubMed]
50. Budimlija, Z.M.; Prinz, M.K.; Zelson-Mundorff, A.; Wiersema, J.; Bartelink, E.; Mackinnon, G.; Nazzaruolo, B.L.; Estacio, S.M.; Hennessey, M.J.; Shaler, R.C. World Trade Center Human Identification Project: Experiences With Individual Body Identification Cases. *Croat. Med. J.* **2003**, *44*, 259–263. [PubMed]
51. Galloway, A. (Ed.) *Broken Bones: Anthropological Analysis of Blunt Force Cranial Trauma*; Charles C. Thomas: Springfield, IL, USA, 1999.
52. Schmidt, C.; Symes, S.A. (Eds.) *The Analysis of Burned Human Remains*; Elsevier: London, UK, 2008.
53. Symes, S.A.; L'Abbé, E.N.; Chapman, E.N.; Wolff, I.; Dirkmaat, D.C. Interpreting Traumatic Injury to Bone in Medicolegal Investigations. In *A Companion to Forensic Anthropology*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2012; pp. 340–389.
54. Schotsmans, E.M.; Márquez-Grant, N.; Forbes, S.L. *Taphonomy of Human Remains: Forensic Analysis of the Dead and the Depositional Environment*; John Wiley & Sons: Hoboken, NJ, USA, 2017.

55. Passalacqua, N.V.; Rainwater, C.W. (Eds.) *Skeletal Trauma Analysis: Case Studies in Context*; Wiley Blackwell: Oxford, UK, 2015.
56. Haglund, W.D.; Sorg, M.H. (Eds.) *Advances in Forensic Taphonomy*; CRC Press: Boca Raton, FL, USA, 2002; p. 507.
57. Pokines, J.; Symes, S.A. (Eds.) *Manual of Forensic Taphonomy*; CRC Press: Boca Raton, FL, USA, 2013.
58. Beatrice, J.S.; Soler, A. Skeletal Indicators of Stress: A Component of The Biocultural Profile of Undocumented Migrants In Southern Arizona. *J. Forensic Sci.* **2016**, *61*, 1164–1172. [CrossRef]
59. Soler, A.; Beatrice, J.S. Expanding The Role of Forensic Anthropology In a Humanitarian Crisis: An Example from the USA-Mexico Border. In *Sociopolitics of Migrant Death and Repatriation*; Latham, K.A., O'Daniel, A., Eds.; Springer: Cham, Switzerland, 2018; pp. 115–128.
60. Kimmerle, E.H.; Baraybar, J.P. (Eds.) *Skeletal Trauma: Identification of Injuries Resulting from Human Rights Abuse and Armed Conflict*; CRC Press: Boca Raton, FL, USA, 2008.
61. Tuller, H.H. Mass Graves and Human Rights: Latest Developments, Methods, and Lessons Learned. In *A Companion to Forensic Anthropology*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2012; pp. 157–174.
62. Ubelaker, D. Foreword. In *Forensic Anthropology: An Introduction*; Tersigni-Tarrant, M., Shirley, N.R., Eds.; CRC Press: Boca Raton, FL, USA, 2013; p. xi.
63. Collins, H. Studies of Expertise and Experience. *Topoi* **2018**, *37*, 67–77. [CrossRef]
64. Mody, C.; Kaiser, D. Scientific Training and The Creation of Scientific Knowledge. In *The Handbook of Science and Technology Studies*; Hackett, E.J., Amsterdamska, O., Lynch, M.J., Wajcman, J., Eds.; MIT Press: Cambridge, MA, USA, 2008; pp. 377–402.
65. Bourdieu, P.; Passeron, J.-C. *Reproduction in Education, Society and Culture*; Sage: London, UK, 1977.
66. Bourdieu, P. *Homo Academicus*; Stanford University Press: Palo Alto, CA, USA, 1988.
67. Spring, J. *The Sorting Machine Revisited: National Educational Policy Since 1945*; Longman: New York, NY, USA, 1989.
68. Kliebard, H. *Schooled to Work: Vocationalism and The American Curriculym, 1876–1946*; Teachers College Press: New York, NY, USA, 1999.
69. Snow, C.C. Forensic Anthropology. *Annu. Rev. Anthrpol.* **1982**, *11*, 97–131. [CrossRef]
70. Tersigni-Tarrant, M.; Shirley, N.R. (Eds.) Brief History of Forensic Anthropology. In *Forensic Anthropology: An Introduction*; CRC Press: Boca Raton, FL, USA, 2013; pp. 1–16.
71. Buikstra, J.E.; Ubelaker, D.H. (Eds.) *Standards for Data Collection from Human Skeletal Remains*; Arkansas Archeological Survey Research Series No. 44: Fayetteville, AR, USA, 1994.
72. Larsen, C.S. *Bioarchaeology: Interpreting Behavior from the Human Skeleton*; Cambridge University Press: Cambridge, UK, 1997.
73. Berryman, H.E.; Lanfear, A.K. Forensic Anthropologists In Medical Examiner's and Coroner's Offices: A History. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; John Wiley & Sons, Ltd.: Malden, MA, USA, 2012; pp. 534–548.
74. Buikstra, J.E.; King, J.L.; Nystrom, K.C. Forensic Anthropology and Bioarchaeology in The American Anthropologist Rare But Exquisite Gems. *Am. Anthropol.* **2003**, *105*, 38–52. [CrossRef]
75. Bethard, J.D. Historical Trends in Graduate Research and Training of Diplomates of The American Board of Forensic Anthropology. *J. Forensic Sci.* **2017**, *62*, 5–11. [CrossRef]
76. Baker, B.J.; Agarwal, S.C. Stronger Together: Advancing A Global Bioarchaeology. *Bioarchaeol. Int.* **2017**, *1*, 1–18. [CrossRef]
77. Passalacqua, N.V.; Christensen, A.M.; Hefner, J.T. Forensic Anthropology: A Journal for Our Discipline. *Forensic Anthropol.* **2018**, *1*, 1–3. [CrossRef]
78. Martin, D.L. Excavating for Truths: Forensic Anthropology and Bioarchaeology As Ways of Making Meaning From Skeletal Evidence. In *Disturbing Bodies: Perspectives on Forensic Anthropology*; Crossland, Z., Joyce, R.A., Eds.; School for Advanced Research Press: Santa Fe, CA, USA, 2015; pp. 157–168.
79. Martin, D.L.; Anderson, C.P. (Eds.) *Bioarchaeological and Forensic Perspectives on Violence: How Violent Death Is Interpreted from Skeletal Remains*; Cambridge University Press: Cambridge, UK, 2014.
80. Anderson, C.P.; Martin, D.L. (Eds.) *Massacres: Bioarchaeology and Forensic Anthropology Approaches*; University Press of Florida: Gainesville, FL, USA, 2018.
81. Scott, D.D.; Connor, M. The Role and Future of Archaeology in Forensic Science. *Hist. Archaeol.* **2001**, *35*, 101–104. [CrossRef]
82. Skinner, M.; Alempijevic, D.; Djuric-Srejic, M. Guidelines for International Forensic Bio-Archaeology Monitors of Mass Grave Exhumations. *Forensic Sci. Int.* **2003**, *134*, 81–92. [CrossRef]
83. Jessee, E.; Skinner, M. A Typology of Mass Grave and Mass Grave-Related Sites. *Forensic Sci. Int.* **2005**, *152*, 55–59. [CrossRef] [PubMed]
84. Steadman, D.W. The Body-As-Evidence Paradigm In Domestic and International Forensic Anthropology. In *Bioarchaeologists Speak Out: Bioarchaeology and Social Theory*; Buikstra, J.E., Ed.; Springer: Cham, Switzerland, 2019; pp. 243–255.
85. Passalacqua, N.V.; Pilloud, M.A. Education and Training In Forensic Anthropology. *Forensic Anthropol.* **2020**, *3*, 65–74. [CrossRef]
86. Evans, R.; Collins, H. Expertise: From Attribute to Attribution and Back Again? In *The Handbook of Science and Technology Studies*; Hackett, E., Amsterdamska, O., Lynch, M., Wajcman, J., Eds.; MIT Press: Cambridge, MA, USA, 2008; pp. 609–630.
87. Nichols, T. *The Death of Expertise: The Campaign Against Established Knowledge and Why It Matters*; Oxford University Press: New York, NY, USA, 2017.
88. Collins, H.M.; Evans, R. The Third Wave of Science Studies: Studies of Expertise and Experience. *Soc. Stud. Sci.* **2002**, *32*, 235–296. [CrossRef]
89. Dreyfus, H.L.; Dreyfus, S.E. *Mind Over Machine: The Power of Human Intuition and Expertise in the Age of the Computer*; Basil Blackwell: Oxford, UK, 1986.
90. Nowotny, H. Transgressive Competence: The Narrative of Expertise. *Eur. J. Soc. Theory* **2000**, *3*, 5–21. [CrossRef]

91. Gorman, M.E. Levels of Expertise and Trading Zones: A Framework for Multidisciplinary Collaboration. *Soc. Stud. Sci.* **2002**, *32*, 933–938. [CrossRef]
92. Nowotny, H. Dilemma of Expertise. *Sci. Public Policy* **2003**, *30*, 151–156. [CrossRef]
93. Collins, H. Interactional Expertise as a Third Kind of Knowledge. *Phenomenol. Cogn. Sci.* **2004**, *3*, 125–143. [CrossRef]
94. Collins, H. How Do You Know You've Alternated? *Soc. Stud. Sci.* **2004**, *34*, 103–106. [CrossRef]
95. Selinger, E.; Mix, J. On Interactional Expertise: Pragmatic and Ontological Considerations. *Phenomenol. Cogn. Sci.* **2004**, *3*, 145–163. [CrossRef]
96. Carolan, M.S. Science, Expertise, and The Democratization of The Decision-Making Process. *Soc. Nat. Resour.* **2006**, *19*, 661–668. [CrossRef]
97. Evans, R. The Sociology of Expertise: The Distribution of Social Fluency. *Sociol. Compass* **2008**, *2*, 281–298. [CrossRef]
98. Evans, R.; Collins, H. Interactional Expertise and The Imitation Game. In *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*; Gorman, M.E., Ed.; MIT Press: Cambridge, MA, USA, 2010; pp. 53–70.
99. Dasgupta, D.; Levine, A. Trading Zones, Peripheral Science and Contributory Expertise: A Framework for Exploring Science Outside The Western Contexts. *Curr. Sci.* **2013**, *104*, 1616–1618.
100. Collins, H. Three Dimensions of Expertise. *Phenomenol. Cogn. Sci.* **2013**, *12*, 253–273. [CrossRef]
101. Goddiksen, M. Clarifying Interactional and Contributory Expertise. *Stud. Hist. Philos. Sci. Part A* **2014**, *47*, 111–117. [CrossRef]
102. Collins, H.; Evans, R. Expertise Revisited, Part I—Interactional Expertise. *Stud. Hist. Philos. Sci. Part A* **2015**, *54*, 113–123. [CrossRef] [PubMed]
103. Nimkulrat, N.; Niedderer, K.; Evans, M. On Understanding Expertise, Connoisseurship, and Experiential Knowledge in Professional Practice. *J. Res. Pract.* **2015**, *11*, 1–13.
104. Reyes-Galindo, L.I.; Duarte, T.R. Bringing Tacit Knowledge Back to Contributory and Interactional Expertise: A Reply to Goddiksen. *Stud. Hist. Philos. Sci. Part A* **2015**, *49*, 99–102. [CrossRef]
105. Collins, H.; Evans, R.; Weinel, M. Expertise Revisited, Part II: Contributory Expertise. *Stud. Hist. Philos. Sci. Part A* **2016**, *56*, 103–110. [CrossRef] [PubMed]
106. Ribeiro, R.; Lima, F.P. The Value of Practice: A Critique of Interactional Expertise. *Soc. Stud. Sci.* **2016**, *46*, 282–311. [CrossRef] [PubMed]
107. Collins, H.; Evans, R.; Weinel, M.; Lyttleton-Smith, J.; Bartlett, A.; Hall, M. The Imitation Game and The Nature of Mixed Methods. *J. Mix. Methods Res.* **2017**, *11*, 510–527. [CrossRef]
108. Collins, H.; Evans, R. *Rethinking Expertise.*; University of Chicago Press: Chicago, IL, USA, 2007.
109. Dunning, D. The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance. *Adv. Exp. Soc. Psychol.* **2011**, *44*, 247–296.
110. Huang, S. When Peers Are Not Peers and Don't Know It: The Dunning-Kruger Effect and Self-Fulfilling Prophecy in Peer-Review. *Bioessays* **2013**, *35*, 414–416. [CrossRef] [PubMed]
111. Kruger, J.; Dunning, D. Unskilled and Unaware of It: How Difficulties In Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *J. Pers. Soc. Psychol.* **1999**, *77*, 1121–1134. [CrossRef]
112. Pennycook, G.; Ross, R.M.; Koehler, D.J.; Fugelsang, J.A. Dunning–Kruger Effects In Reasoning: Theoretical Implications of the Failure to Recognize Incompetence. *Psychon. Bull. Rev.* **2017**, *24*, 1774–1784. [CrossRef]
113. Passalacqua, N.V.; Pilloud, M.A. *Ethics and Professionalism in Forensic Anthropology*; Academic Press: San Diego, CA, USA, 2018.
114. Dewitte, S.N. Bioarchaeology and The Ethics of Research Using Human Skeletal Remains. *Hist. Compass* **2015**, *13*, 10–19. [CrossRef]
115. Lambert, P.M.; Walker, P.L. Bioarchaeological Ethics: Perspectives tn the Use and Value of Human Remains in Scientific Research. In *Biological Anthropology of the Human Skeleton*; Katzenberg, A., Grauer, A.L., Eds.; Wiley Blackwell: Hoboken, NJ, USA, 2019; pp. 3–42.
116. Scaramelli, F.; Scaramelli, K. Archaeology, Bioethics, and Policies Regarding The Treatment of Ancient Human Remains In Venezuela. In *Archaeological Human Remains*; O'Donnabhain, B., Lozada, M.C., Eds.; Springer: New York, NY, USA, 2014; pp. 235–245.
117. Turner, T.R. (Ed.) Introduction: Ethical Concerns in Biological Anthropology. In *Biological Anthropology and Ethics: From Repatriation to Genetic Identity*; State University of New York Press: Albany, NY, USA, 2005; pp. 1–14
118. Turner, T.R.; Wagner, J.K.; Cabana, G.S. Ethics in Biological Anthropology. *Am. J. Phys. Anthropol.* **2018**, *165*, 939–951. [CrossRef]
119. Walker, P.L. Bioarchaeological Ethics: A Historical Perpective on The Value of Human Remains. In *Biological Anthropology of The Human Skeleton*; Katzenberg, A., Saunders, S., Eds.; Wily-Liss: Wilimington, NC, USA, 2008; pp. 3–39.
120. Blau, S. More Than Just Bare Bones: Ethical Considerations for Forensic Anthropologists. In *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S., Ubelaker, D., Eds.; Routledge: New York, NY, USA, 2016; pp. 593–606.
121. Cordner, S.; Leahy, F. Ethics of Research On Bodies In The Jurisdiction of Coroners Or Medical Examiners. *Acad. Forensic Pathol.* **2013**, *3*. [CrossRef]
122. France, D.L. Ethics in Forensic Anthropology. In *A Companion to Forensic Anthropology*; Dirkmaat, D., Ed.; John Wiley & Sons, Ltd.: Malden, MA, USA, 2012; pp. 666–682.
123. Márquez-Grant, N.; Passalacqua, N.V.; Pilloud, M.A.; Lester, N.; Decker, S.; Ford, J. Ethical Concerns in Forensic Anthropology. In *Ethical Approaches to Human Remains: A Global Challenge in Bioarchaeology and Forensic Anthropology*; Squires, K., Errickson, D., Marquez-Grant, N., Eds.; Springer: Cham, Switzerland, 2020; pp. 347–366.
124. Walsh-Haney, H.A.; Lieberman, L.S. Ethical Concerns in Forensic Anthropology. In *Biological Anthropology and Ethics: From Repatriation to Genetic Identity*; Turner, T., Ed.; State University of New York Press: Albany, NY, USA, 2005; pp. 121–132.

125. Dyrbye, L.N.; Massie, F.S.; Eacker, A.; Harper, W.; Power, D.; Durning, S.J.; Thomas, M.R.; Moutier, C.; Satele, D.; Sloan, J. Relationship Between Burnout and Professional Conduct and Attitudes Among US Medical Students. *J. Am. Med. Assoc.* **2010**, *304*, 1173–1180. [CrossRef]
126. Harris, C.E. Engineering Ethics: From Preventive Ethics to Aspirational Ethics. In *Philosophy and Engineering: Reflections on Practice, Principles and Process*; Michelfelder, D., Mccarthy, N., Goldberg, D., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 177–187.
127. Appelbaum, P.S. A Theory of Ethics for Forensic Psychiatry. *J. Am. Acad. Psychiatry Law* **1997**, *25*, 233–248. [PubMed]
128. Sun, L. How America Hates Socialism without Knowing Why. Available online: https://evonomics.com/america-hates-socialism-without-knowing/ (accessed on 16 July 2021).
129. Lizza, R. How Trump Broke The Office of Government Ethics. New Yorker. Available online: https://www.newyorker.com/news/ryan-lizza/how-trump-broke-the-office-of-government-ethics (accessed on 14 July 2017).
130. Mcgehee, M.; Gray, W.; Simmons, C.; Horvath, M. The Ethics Blind Spot: How the House and Senate Ethics Committees Fail to Uphold High Ethical Standards-and Solutions to Fix the Problem. 2018. Available online: https://www.Issueone.Org/Wp-Content/Uploads/2018/02/Ethics-Blind-Spot-Final.Pdf (accessed on 16 July 2021).
131. Josephson, M. *Making Ethical Decisions*; Josephson Institute of Ethics: Los Angeles, CA, USA, 2002.
132. Galloway, A.; Simmons, T. Education in Forensic Anthropology: Appraisal and Outlook. *J. Forensic Sci.* **1997**, *42*, 796–801. [CrossRef]
133. Swganth. Qualifications. 2010. Available online: https://www.Nist.Gov/System/Files/Documents/2018/03/13/Swganth_Qualifications.Pdf (accessed on 16 July 2021).
134. Swganth. Education and Training. 2013. Available online: https://www.Nist.Gov/System/Files/Documents/2018/03/13/Swganth_Education_And_Training.Pdf (accessed on 16 July 2021).
135. Langley, N.; Tersigni-Tarrant, M. Core Competencies In Foernsic Anthropology: A Framework for Education, Training, and Practice. *Forensic Anthropol.* **2020**, *3*, 75–80. [CrossRef]
136. Crowther, W. Native American Graves Protection and Repatriation Act: How Kennewick Man Uncovered the Problems in NAGPRA. *J. Land Resour. Environ. Law* **2000**, *20*, 269–290.
137. Zimmerman, L.J. Public Heritage, a Desire for a White History for America, and Some Impacts of the Kennewick Man/Ancient One Decision. *Int. J. Cult. Prop.* **2005**, *12*, 265–274. [CrossRef]
138. Nicholas, G.P.; Jules, J.; Dan, C.; Zimmerman, L.; Smith, C. Moving Beyond Kennewick: Other Native American Perspectives On Bioarchaeological Data and Intellectual Property Rights. In *Kennewick Man: Perspectives on the Ancient One*; Burke, H., Smith, C., Lippert, D., Watkins, J., Zimmerman, L., Eds.; Routledge: London, UK, 2008; pp. 233–243.
139. Passalacqua, N.V.; Pilloud, M.; Belcher, W.R. Scientific Integrity In The Forensic Sciences: Consumerism, Conflicts of Interest, and Transparency. *Sci. Justice* **2019**, *59*, 573–579. [CrossRef] [PubMed]
140. Rosenblatt, A. International Forensic Investigations and The Human Rights of The Dead. *Hum. Rights Q.* **2010**, *32*, 921–950. [CrossRef]
141. Kramer, M.H. Do Animals and Dead People Have Legal Rights. *Can. J. Law Jurisprud.* **2001**, *14*, 29–54. [CrossRef]
142. Moon, C. What Remains? Human Rights after Death. In *Ethical Approaches to Human Remains*; Squires, K., Errickson, D., Marquez-Grant, N., Eds.; Springer: Cham, Switzerland, 2019; pp. 39–58.
143. Shui, B. Where Law Meets Culture: The Legal Protection of The Dead in China. *Int. Comp. Law Rev.* **2015**, *23*, 139–192.
144. Pryzwansky, W. The Regulation of School Psychology: A Historical Perspective on Certification, Licensure, and Accreditation. *J. Sch. Psychol.* **1993**, *31*, 219–235. [CrossRef]
145. SWGTOX. Standard for Laboratory Personnel. *J. Anal. Toxicol.* **2015**, *39*, 241–250.

biology

MDPI

Article

Ancestry Studies in Forensic Anthropology: Back on the Frontier of Racism

Ann H. Ross [1,*] and Shanna E. Williams [2]

[1] Human Identification & Forensic Analysis Laboratory, Department of Biological Sciences, North Carolina State University, Raleigh, NC 276995, USA
[2] Department of Biomedical Sciences, UofSC School of Medicine Greenville, Greenville, SC 29605, USA; WILL3992@greenvillemed.sc.edu
* Correspondence: ahross@ncsu.edu

Simple Summary: Within the practice of forensic anthropology ancestry is oftentimes used as a proxy for social race. This concept and its implications were explored via a content analysis (2009–2019) of the Journal of Forensic Sciences. Our findings revealed antiquated views of race based on the trifecta of continental populations (Asia, Europe, and Africa) continue to be pervasive in the field despite scientific invalidation of the concept of race decades earlier. Moreover, our employment of modern geometric morphometric and spatial analysis methods on craniofacial coordinate anatomical landmarks from several Latin American samples produced results in which the groups were not patterned by ancestry trifecta. Based on our findings we propose replacing the assumption of continental ancestry with a population structure approach that combines microevolutionary and cultural factors with historical events in the examination of population affinity.

Abstract: One of the parameters forensic anthropologists have traditionally estimated is ancestry, which is used in the United States as a proxy for social race. Its use is controversial because the biological race concept was debunked by scientists decades ago. However, many forensic anthropologists contend, in part, that because social race categories used by law enforcement can be predicted by cranial variation, ancestry remains a necessary parameter for estimation. Here, we use content analysis of the *Journal of Forensic Sciences* for the period 2009–2019 to demonstrate the use of various nomenclature and resultant confusion in ancestry estimation studies, and as a mechanism to discuss how forensic anthropologists have eschewed a human variation approach to studying human morphological differences in favor of a simplistic and debunked typological one. Further, we employ modern geometric morphometric and spatial analysis methods on craniofacial coordinate anatomical landmarks from several Latin American samples to test the validity of applying the antiquated tri-continental approach to ancestry (i.e., African, Asian, European). Our results indicate groups are not patterned by the ancestry trifecta. These findings illustrate the benefit and necessity of embracing studies that employ population structure models to better understand human variation and the historical factors that have influenced it.

Keywords: race; ancestry; population affinity; craniofacial variation; geometric morphometrics

Citation: Ross, A.H.; Williams, S.E. Ancestry Studies in Forensic Anthropology: Back on the Frontier of Racism. *Biology* **2021**, *10*, 602. https://doi.org/10.3390/biology10070602

Academic Editor: Andrés Moya

Received: 18 May 2021
Accepted: 28 June 2021
Published: 29 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Forensic anthropology is a sub-discipline of biological anthropology, the science of studying what it means to be human via our biology. Forensic anthropologists are experts in human skeletal anatomy, growth, and development; expertise that we use in medicolegal death investigations for the recovery and analysis of human skeletal remains. A significant part of our analysis is the creation of the biological profile, an evaluation of four criteria that may assist with identification: age-at-death, sex (for adult skeletons), stature, and ancestry [1]. The estimation of ancestry is one of the most difficult (and controversial) parameters of the biological profile. It is often conflated with social race and ethnicity

by medical examiners, law enforcement, forensic practitioners, and government agencies. Further, some practitioners have questioned the validity of estimating this parameter and if the estimation could even hinder identification because of racial bias on the part of investigative agencies [2–4]. Part of the reason its use is so controversial is that the biological race concept, namely, that the human species can be divided into biological races, was debunked decades ago [5]. In the 1990s there was discord within biological anthropology stimulated by a paper by Lieberman and colleagues [6], presented earlier in 1987 at the American Association of Physical Anthropologists annual meeting that reported 50% of the biological anthropologists polled believed in the race concept. Forensic anthropologists argued that it was a pragmatic decision to include "race" in their forensic case reports as "race" was used by law enforcement and medicolegal death investigators working the missing and unidentified cases [7]. Thus, in 1992 a name change from "race" to "ancestry" was proposed as a less loaded term [7]. This has been rationalized by the notion that we can connect craniofacial morphology (i.e., size and shape variants of skull bone features) to social race categories (e.g., United States Census categories) [8,9]. However, some biological anthropologists questioned the ethics of even estimating this parameter fearing that its continued use would endorse racist views and be complicit in the social injustices faced by underrepresented groups [2,10–12].

In a search for the term "ancestry" in the titles of the *Journal Forensic Sciences (JFS)* between the years of 2009–2019, 20 articles used "ancestry" and in 2010 and 2011, two articles still used "race." The term ancestry appeared 24 times in the keywords between 2009–2019, with four papers using samples identified as *black*, *white*, and *Hispanic*. Five papers used samples identified as *black* and *white*, which included a paper on *South African blacks* and *whites*. There were 12 papers with various iterations of "*Hispanic*" (i.e., *South West Hispanics*); as well as papers that defined their samples as *Prehistoric Native Americans*; those that use a few country names (e.g., Japanese, Guatemala, Germany, Thailand, etc.); and a paper on *Native American*, *Japanese*, and *Thai* samples. This literature review clearly illustrates the lack of purpose, consensus, and consistent usage of the nomenclature; suggesting that the transition from race to ancestry was primarily a linguistic change (see [13] that covers the problems with nomenclature). The many iterations of "Hispanic" are a result of the 2008 migrant death symposium at the American Academy of Forensic Sciences annual meeting dealing with the difficulty of identifying unidentified border-crossers (UBCs) in the United States. Interestingly, the term Hispanic is still commonly used even though it has no biological meaning [14], and going as far back as 1992, pioneering forensic anthropologist Alice Brues understood that "Hispanics" from South Florida, the Southwest, and Texas should not be grouped under one umbrella because they represented different population migrations to the US [14,15].

The results of this literature review also illustrate the return to antiquated and over-simplistic views of race based on the trifecta of continental populations from Asia, Europe, and Africa used by typologists of the early 20th century, have regained popularity [16]. In part, this is because the reference databases we rely on to compare cranial measurements of an unknown person were constructed using such categories. However, this facile presumption ignores underlying microevolutionary mechanisms such as drift, migrations, and mutation that are responsible for human variation and diversity. Studies of global populations reveal that human craniofacial morphology fits a neutral evolutionary model because contiguous populations more frequently exchange genes and/or share common ancestry [17].

Therefore, rather than studying population affinity via an assumption of continental ancestry, we instead advocate for a population structure approach. The benefit of such an approach is that it allows us to understand how microevolutionary factors such as genetic drift act in concert with cultural factors (i.e., marriage patterns) and historical events (i.e., epidemics, colonization) to influence human variation. A population structure approach is empirically driven, meaning that it is based on firm observations without phylogenetic assumptions and by operational approaches that are hypothesis-driven by

meaningful questions [18]. When comparing populations one can select various types of characters for investigation such as morphology, physiology, behavior, and/or ecology. However, common mistakes made in the selection of a character for estimating similarity is a failure to identify the biological factors that the characters represent (i.e., their heredity) and assumptions that they are all equally informative in providing evidence of group (i.e., phenetic) similarity [18]. One example of the former is with the use of the skull trait variant post-bregmatic depression [3,4]. As noted, a major consideration in the application of a population structure approach is to account for historical events such as population influxes and settlements, religious secularization, language differences, temporality, and spatial patterning that would be impacted by microevolutionary forces [19].

Myopically, forensic anthropology abandoned the study of human biological variation based on a strong foundation of examining human variation through a population structure lens grounded in microevolution, and instead re-embraced a typological approach that looks a lot like "race" of the early years [20,21]. Therefore, it is clear that a broad synthesis to better understand the underlying patterns of modern human variation that would disclose the underlying population structure of the group(s) under study is needed. Such information would also be of use to biological anthropology more broadly. Here, we use craniofacial coordinate anatomical landmarks from Latin American samples while employing modern geometric morphometric and spatial analysis methods to test the validity of applying the antiquated tri-continental approach to ancestry. These samples were chosen given the stated problems with the comprehensive, non-critical use of the "*Hispanic*" label for anyone from Latin America or Spain, and in an attempt to partition out how different historical socio-political events within Latin America have influenced biological variation. Further, we discuss how situating such approaches within a microevolutionary framework can enrich our understanding of how major historical events influence human variation and population structure.

2. Materials and Methods

2.1. Samples

The sample totals 397 modern adult individuals and includes individuals from Latin America (Chile, Colombia, Cuba, Guatemala, Panama, Puerto Rico, and Peru); and comparative skeletal samples from Spain and enslaved Africans from Cuba were included to explore the effects of colonialism and the Transatlantic Slave Trade on the population structure of the region. Males and females were analyzed separately when this information was available (see Table 1). Some samples were small due to poor preservation in tropical environments. To incorporate all of the observed biological information and to increase sample sizes males and females were pooled as it has been found that sex variation is negligible within each population included in population [22]. Latitude and longitude were recorded based on present-day political boundaries. The sample composition is presented in Table 1.

While we acknowledge the value data collected from such samples continue to contribute to discussions of human variation, it should also be noted that the history and ethics of human skeletal collections, in general, is often dubious. Such body harvesting all too often occurred under the umbrella of scientific racism, without the permission of the deceased or next of kin, and disproportionately targeted marginalized populations.

Sixteen type 1 and 2 standard anatomical craniofacial landmarks (for a total number of landmarks 16 $\times$ 3 dimensions = 48) that should reflect the among-group variation were utilized in the analyses (Table 2 and Figure 1). The landmarks selected were those that are of particular interest in forensic anthropology and that would allow for broader shape coverage. To mitigate the effect of small sample sizes, a PCA was used as a dimension-reducing technique and limiting the number of variables [23,24].

Table 1. Sample composition and provenience.

Group	N	Provenance	Latitude	Longitude
Chile	♀= 34 ♂= 37	Juan Munizaga Collection, Universidad de Chile, Santiago, Chile	−33.45	−70.67
Colombia	♀= 11 ♂= 53	Antioquia modern skeletal collection, Escuela Nacional de Criminalística, Medellín, Colombia	6.230833	−75.5906
Cuba	19	Cemetery Collection, Museo de Montane, Havana, Cuba	23.11359	−82.3666
Enslaved Africans	25	Morton Collection, University of Pennsylvania, US	−8.83833	13.23444
Guatemala	♂= 71	Provided by Kate Spradley	14.62843	−90.5227
Puerto Rico	♂= 5	University of Rio Piedras, Puerto Rico	18.46633	−66.1057
Panama	10	Insituto de Medicina Legal, Panama	8.983333	−79.5167
Peru	7	C.A. Pound Human ID Lab, University of Florida, US	−12.0464	−77.0428
Spain	♀= 58 ♂= 67	Oloriz Collection, Madrid, Spain	40.41678	−3.70379

Table 2. Anatomical landmarks and associated numbers.

Landmark Number	Anatomical Landmarks
1, 2	Asterion, bilateral
3	Bregma
4, 5	Dacryon, bilateral
6, 7	Ectoconchion, bilateral
8,9	Frontomalare temporale, bilateral
10	Lambda
11	Nasion
12	Subspinale
13, 14	Zygomaxillare, bilateral
15, 16	Zygoorbitale, bilateral

2.2. Landmark Precision and Reliability

Only type 1 and type 2 landmarks were included as they have been found to be reliably reproducible [25]. The landmarks included are those that were found to meet the less than 5 percent error threshold for digitizing and intra-observer error [25]. The coordinate data were collected using a Microscribe G2X digitizer with a reported average error rate of 239 mm [26]. These samples are part of the reference database for the classification software 3D-ID [27] and prior to inclusion in the software, data underwent extensive error checks via mapping (i.e., visualization) of all individuals using the Generalized Procrustes analysis or GPA function in Morpheus et al. [28].

2.3. Geometric Morphometrics

Before statistical analyses can be performed, coordinate data must first undergo a GPA transformation using the software *MorphoJ*, which is freely available for downloading and developed by Klingenberg [29]. GPA brings all specimens into a common coordinate system, after it translates, rotates, and scales each individual. The advantage of this method is that morphological shape and size can be examined separately, with shape defined as all of the geometric information that remains after the effects of location, scale, and rotational effects are removed [30,31]. Centroid size is defined as a measure of geometric scale that is mathematically independent of shape [31]. To reduce the dimensionality, a principal component analysis (PCA) of the covariance matrix was performed on the GPA-transformed

coordinate data and these principal components were utilized for ensuing statistical analyses [31]. A canonical variates analysis (CVA) was performed to examine the most amount of the variation with the least dimensions possible of the a priori groups [29]. A generalized distance measure (or Mahalanobis distance) was used to examine group similarity [29]. A discriminant function analysis (DFA) was performed to visualize morphological variation between the consensus configurations of each group. The phenetic (e.g., morphological) among-group variation was examined using ANOVA for centroid size. Among-group variation for shape was analyzed using MANOVA of the principal components scores derived from MorphoJ. The ANOVA and MANOVA procedures were performed in JMP® Pro 14.1 [32].

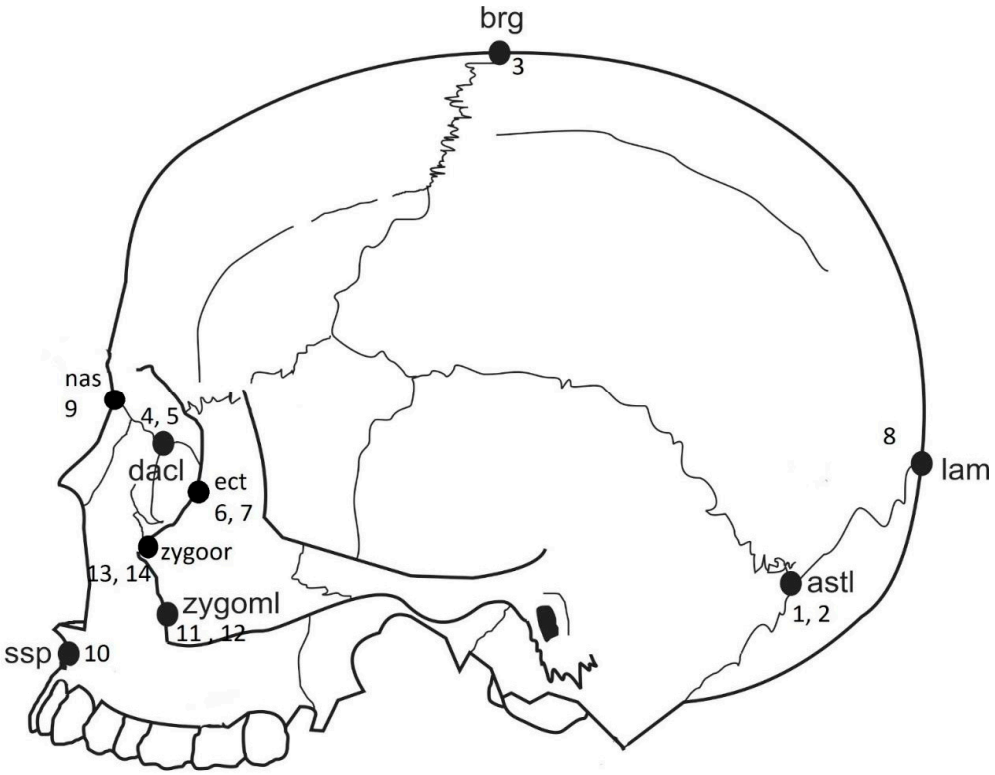


Figure 1. Anatomical landmark location and associated landmark number from Table 1.

2.4. Hierarchical Clustering

Average linkage hierarchical (or agglomerative) clustering was conducted using the generalized distance matrix to examine group similarity [33,34]. The process begins with each population sample in a single cluster, then in each successive iteration, it merges the closest pair of clusters until all the data is in one cluster. The cluster analysis was performed in JMP® Pro 14.1 [32].

2.5. Spatial Analysis

Moran's *I*, a product-moment coefficient, was used to measure the spatial autocorrelation of shape (PC1 as only one variable can be utilized) and centroid size, which is a measure of genetic similarity between individuals with reference to geographic separation (latitude/longitude). Spatial correlograms were computed to evaluate the spatial autocorrelation coefficients for all pairs of localities at specified geographic distance classes [35], and were performed using the freeware software GeoDa v1.14.0 [36].

3. Results

3.1. Geometric Morphometrics

Forty-one PC scores were generated from the covariance matrix, which were used as new variables in the subsequent statistical analyses. The ANOVA shows that size is significantly different among the groups (Centroid size: ($F(11, 385) = 22.35$, $p \leq 0.0001$). The MANOVA (of 41 principal component scores derived by MorphoJ) also detected significant shape variation (Shape: Wilks' Lambda 0.0058, df = 451, 3706.6, F = 5.12, $p \leq 0.0001$). The anatomical landmarks used here are in the same location on each skull; this property enables evaluation and observation of any distinctions in overall cranial shape and size between groups. Morphological variation is illustrated via wireframe graphs that depict the magnitude and direction of shape change between two mean configurations with the direction of change depicted from light (light blue) to dark (blue). The starting shape is that of one sample mean configuration that is deformed into a target shape (second sample) mean configuration to visualize the differences. The groups illustrated were selected according to the clusters produced by the hierarchical cluster analysis. The similarity between the Chilean male mean configuration (light blue) and the Spanish male mean configuration (blue) is visualized showing little to no variation in the placement of the anatomical landmarks (Figure 2).

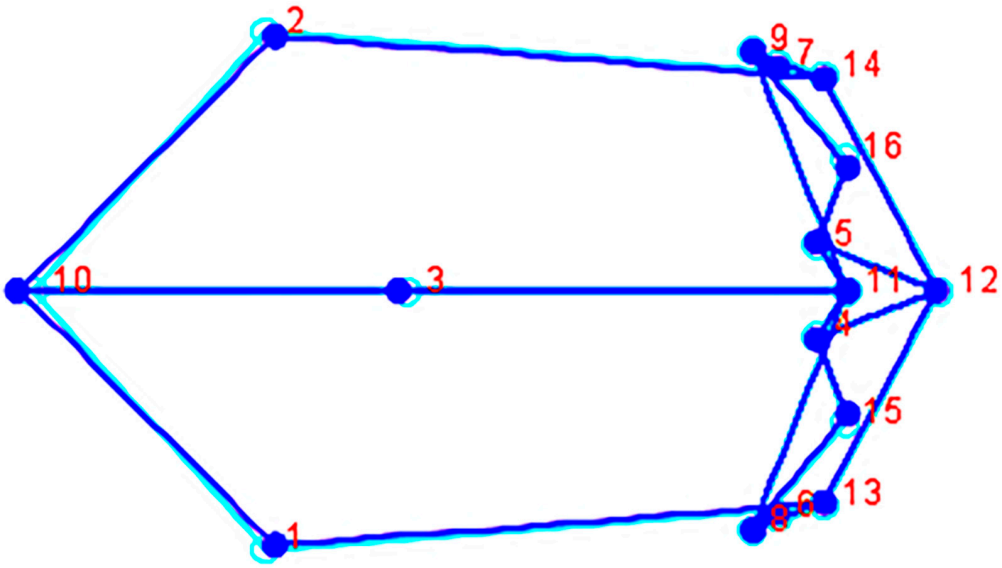


Figure 2. Wireframe (superior view) depicting the Chilean male mean configuration (starting shape, light blue) deformed into the Spanish male mean configuration (target shape, blue). The numbers correspond to the landmarks in Table 2.

To illustrate the importance of a population approach, Panama and Colombia, Panama and enslaved Africans, and Panama and Spanish consensus configurations were compared based on known historical events (i.e., conquest, colonialism, and slavery). The morphological differences between the Colombians and the Panamanians show that the Colombians (light blue) have shorter and narrower crania than Panamanians (blue), depicted by the more posteriorly and inferiorly placed anatomical landmarks bregma and lambda and more superiorly placed anatomical landmarks asterion and zygomaxillare (Figure 3). It also shows that Colombians have a longer upper facial height with the anatomical landmark nasion positioned more superiorly and a more inferiorly placed anatomical landmark zygomaxillare. Enslaved Africans (light blue) have longer and narrower cranial vaults with anatomical landmark lambda more posteriorly placed and asterion more anteriorly placed compared to Panamanians (blue). The wireframe depicting the starting shape of Panamanians (light blue) shows that they have shorter cranial vaults and a shorter and more projecting upper face as evidenced by the more anteriorly placed anatomical landmarks

subspinale, bregma, and lambda, and more inferiorly positioned anatomical landmarks bregma and zygomaxillare than the Spaniards' target shape (blue), see Figure 3.

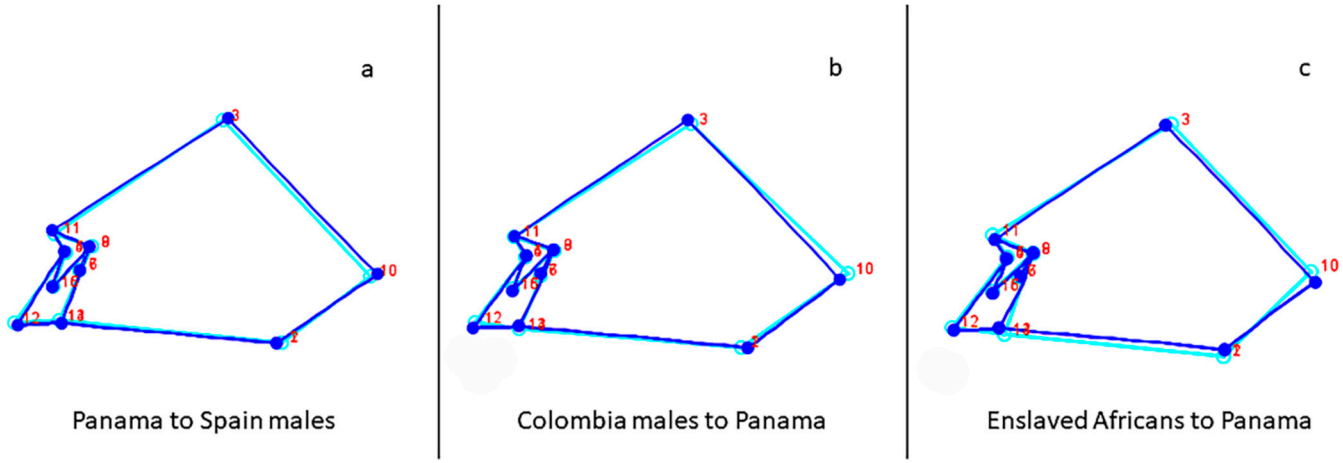


Figure 3. Wireframes depicting the (**a**) Panama (light blue) into Spanish males (blue); (**b**) Colombian males (light blue) into Panama (blue); (**c**) Enslaved Africans (light blue) into Panama (blue). Numbers correspond to landmarks in Table 2.

3.2. Hierarchical Clustering

The dendrogram produced from the hierarchical cluster analysis using the generalized distance matrix shows two distinct clusters: (1) Chile/Spain and (2) Panama, Cuba, Guatemala, and Colombia which branch off the Chile/Spain cluster. The enslaved African sample clusters with Peru, and Puerto Rico is the most dissimilar. This is further illustrated by the constellation plot (Figure 4), which arranges the samples as endpoints. The length of a line between cluster joints represents the distance between them. The plot shows that the most distinct group is the sample from Puerto Rico, which is three times the distance from the Colombian samples and closest to Peru and enslaved African samples. Chileans and Spaniards are closer to each other than to the rest of the groups.

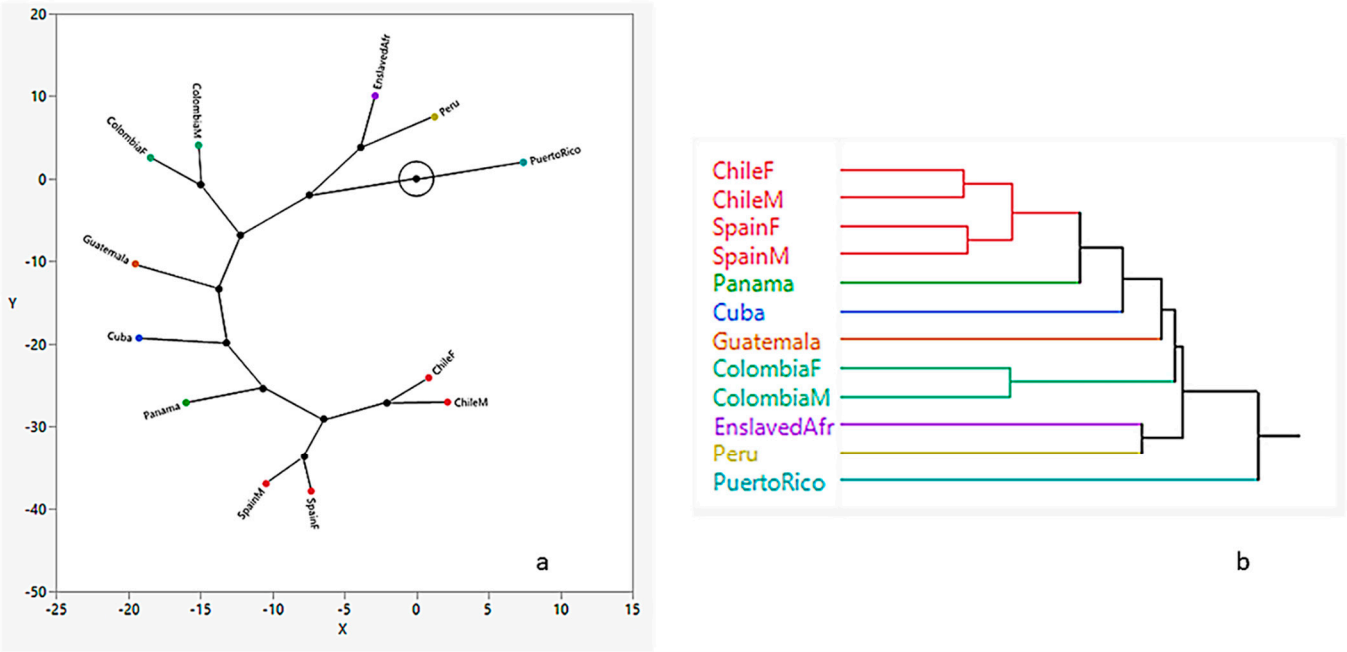


Figure 4. Constellation plot (**a**) and dendrogram (**b**) results from hierarchical cluster analysis showing group relationships.

3.3. Spatial Analysis

The spatial autocorrelation for shape (using PC1 accounting for 21 percent of the total variance) and size show that the groups are spatially patterned and heterogeneous indicated by the positive and significant Z-scores (Table 3). While the correlograms show the autocorrelations decreasing with increased distance, the pattern is generally non-monotonic, meaning the pattern is not clinal as would be expected under an isolation-by-distance model such as kinship [35], for both shape and centroid size. Autocorrelations are expected to be positive at closer distances and negative at greater distances (Figure 5). The correlograms do not support a clinal pattern.

Table 3. Moran's *I* results for shape using the first principal component and size using centroid size with reference to geographic location.

Moran's *I*	Observed	Expected	Std Dev	Z	PR > Z
PC1	0.0695	−0.0025	0.0022	32.5	0.001
CS	0.0027	−0.0025	0.0026	2.0	0.027

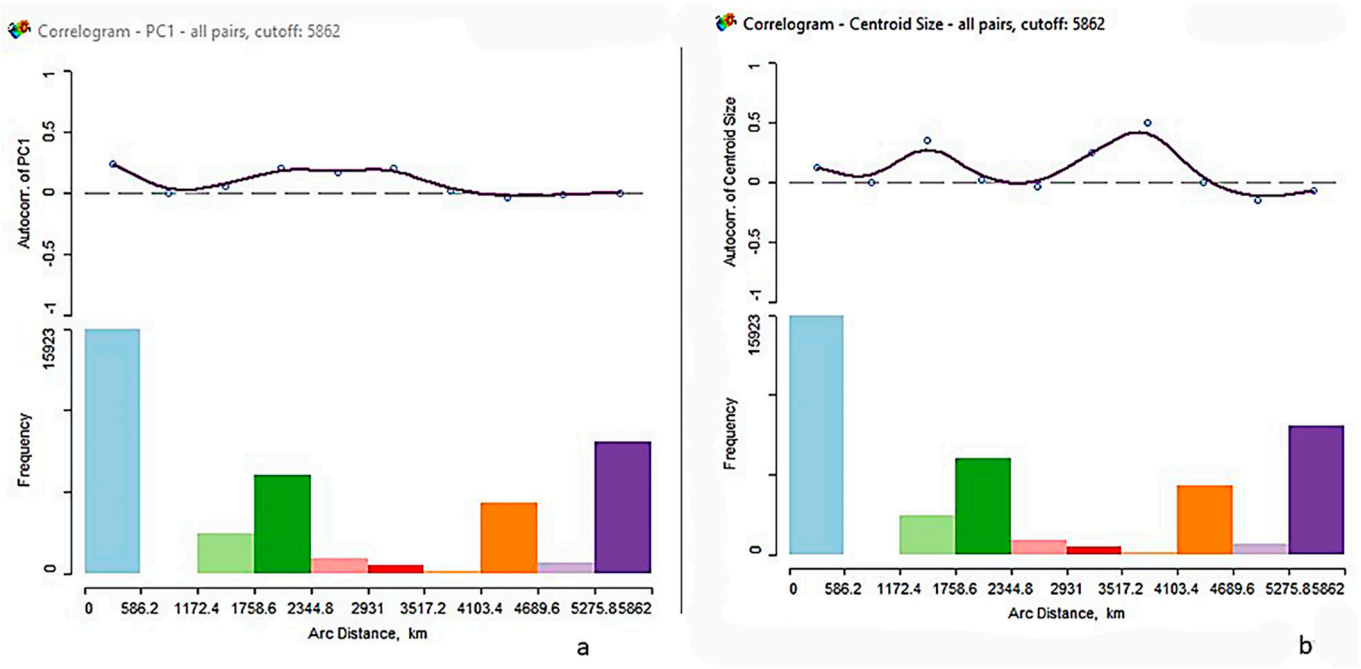


Figure 5. Correlograms for shape (**a**) and size (**b**) depicting the spatial autocorrelation. Moran's *I* by distance in kilometers.

4. Discussion

Even though forensic anthropology as a discipline has moved away from using the term "race" to that of "ancestry", the early critics of race estimation in forensics questioned whether the underlying approach to ancestry would change. Thirty years have passed since this initial criticism and as evidenced by the research published during this time period, ancestry studies have not advanced past the typological (see for example [37]). It is also clear that current research is not fundamentally grounded in an evolutionary framework to understand what has shaped modern human craniofacial [3,4]. The studies surveyed as part of our content analysis show an over-simplistic, typological, tri-continental approach that underscores the need for a paradigm shift to a population structure approach, which incorporates the study of population affinity to understanding modern human biological variation. This paradigm shift can be applied through meaningful hypotheses and avoiding thoughtless comparisons of one sample to another without purpose (e.g., Thai to European Americans, etc.) and by utilizing non-racialized and appropriate reference samples in forensic classification software. For example, implementing nomenclature changes and sample

selection in existing commonly used forensic software such as Fordisc [38], which uses inconsistent terms such as "White, Black, Hispanic, Guatemala, and Japanese", which reflect continental-level, biologically meaningless, and/or country labels; and AncesTrees [39], which uses prehistoric samples that are not applicable for forensic use with antiquated six race categories based in typology, would be a good path forward.

In a recent regional population structure study of pre-contact New World craniofacial variation, Ross and Ubelaker [40] demonstrated that craniofacial variation was a complex interplay between the environment and microevolutionary forces and not the result of a single mechanism. They demonstrated that generally, these pre-Contact populations were spatially patterned, consistent with an isolation-by-distance model. However, they also found a weak association between shape-related variation and altitude, and climate. In the present study, a similar population structure approach was applied to modern Latin American samples to test whether the antiquated trifecta approach to ancestry was valid. Our results demonstrate that Puerto Rico is the most different from the others; Spain and Chile are the most similar to each other compared to the other samples; Panama, Cuba, Guatemala, and Colombia link to the Spain and Chile cluster; and Peru and enslaved Africans form a separate cluster.

The Spanish conquistadors brought enslaved Africans with them beginning as early as 1501 to the Caribbean coast of Panama to colonize the New World [41]. Before the arrival of the Spaniards, there were an estimated 25,000 Amerindians in Panama; by 1522 their population estimates were 13,000 [41]. As a result of the decimation of these Indigenous populations resulting from epidemics and warfare, the Spaniards forced migrations of neighboring Indigenous populations from Panama and Nicaragua; and during Pizarro's expedition to Peru in 1527, 10,000 Amerindians were forcibly displaced to Peru [41]. The association of the Spanish and Chilean samples can be therefore explained through the complex history of conquest and colonialism.

The city of Santiago, Chile was founded in 1542 by Spanish conquistador Pedro de Valdivia. However, the Spanish conquest of Chile was delayed by a long war with Auracanian Indians [42]. During the colonial period, entire Indian populations were decimated by disease and forced labor [42]. From the time of European arrival, slavery of abducted Africans was present, primarily on the Caribbean coast of South America (e.g., Venezuela and Colombia) and in Ecuador and Peru, as well as [42]. Recent work focused on La Isabela, the settlement established after Christopher Columbus' second voyage to what is now the Dominican Republic, suggests that at least one person of African origin was present [43]. The influence of the Transatlantic Slave Trade was detected here by the hierarchical cluster analysis linking Peru and the enslaved African samples. The constellation plot further elucidates the relationship among the groups and illustrates that while the sample from Puerto Rico is the most dissimilar, it is closest to the Peru-enslaved African cluster, followed by Colombia, Guatemala, Cuba, and Panama—all depicting early contact with the Spanish conquistadors that brought enslaved Africans. The spatial analysis was used to assess if there was a spatial pattern based on geographic location. While Moran's *I* was significant and positive for both shape and size, the correlograms show that they are not clinal. The morphological variation for pre-contact populations suggests heterogeneity from the initial population diffusion into the New World prior to European contact [40]. While there is a morphological spatial pattern of modern Latin Americans they do not show a monotonic decrease with distance, but rather indicate repeated population migrations and expansions such as European colonization, the Transatlantic Slave Trade, and forced migrations of Indigenous groups [44]. The argument that there are no races, only clines (or a neutral evolutionary model because neighboring populations more frequently exchange genes and/or share a common ancestry) is not supported here. This finding illustrates a more complex mechanism of modern craniofacial variation and underscores the need for applying a population structure and evolutionary lens to the practice of forensic anthropology.

We use Panama with its complicated history, which has been coveted since the Spanish conquest for its geographic feature as a land bridge of the American continents between the Atlantic and Pacific Oceans, to illustrate the complex nature of assessing population affinity in forensic practice. During the Spanish colonial period, jurisdiction for the Panama territory passed from the Viceroyalty of Lima to Bogotá in the 18th century; it finally gained independence from Spain in 1821 but was part of the Republic of Colombia until 1903 [41]. Importantly, before Panama's split from Colombia, in 1847, a United States merchant set out to build a railroad across the Isthmus that would combine land and sea and open up the Pacific [45,46]. During its construction, a workforce was brought from across the globe (e.g., Austria, China, Colombia, England, France, Germany, India, Ireland, and Jamaica) with thousands dying of malaria, yellow fever, and hardships from the tropical environment [47]. Another important milestone after the failed attempt by the French in the late 1800s was the enormous federally funded undertaking by the United States from 1904–1914 to build an interoceanic canal, a massive earthwork project the likes of which had never been attempted [40,47].

These trans-isthmus ventures brought thousands of migrant workers (~60% from the West Indies) to Panama. The racial contrast of the workers to the engineers and project leaders is crucial to understanding the societal organization and marginalization in the Panama Canal Zone [40]. The colonial caste system transformed into the rigid racial categories imposed by the United States in the Panama Canal Zone, which segregated the workforce both physically and geographically. The Panama Canal Zone was a socialist experiment divided by the white elite minority and the West Indian majority. European Americans showed open disdain for the Panamanians which combined with a culture of flagrant inequality inherited from Spain [40,47]. This segregation, an apartheid not witnessed in any other 20th-century Latin American country [40], was still unmistakable as late as 1986 when the first author graduated from secondary school in the former Zone. Given the complexity of Panama's history, our results are therefore not surprising when viewed against this backdrop. An analysis that rather solely focused on rigid ancestral categories would not have been able to pinpoint Panamanians' dissimilarity to neighboring countries, in particular to Colombia with their shared history under colonial rule. In modern forensic anthropology, all of these heterogeneous groups would have been erroneously designated under the label "Hispanic."

The results of the present study demonstrate that there is substantial diversity in Latin American populations, typically organized into the biologically meaningless grouping of "Hispanic" in contemporary forensic practice. Furthermore, this study obviates the rejection of the tricontinental approach to ancestry estimation and underscores the need for applying a population structure approach with an evolutionary lens to not only understand factors that have influenced craniofacial morphology but test hypotheses about population movements and the impact of major historical events such as conquest and slavery.

5. Conclusions

In 2000, Smay and Armelagos [2] stated that "it was interesting that the word race was being replaced by the less provocative term ancestry", while also indicating they doubted that the logic behind race would change and that the analysis of races using exclusive categories based on folk taxonomy would continue simply under a different name—they were right. Ancestry has become a synonym for race. Given our current global political climate, continuing to type individuals in this way lends credence to existing power structures and socioeconomic inequalities. A mere word change is like putting lipstick on a pig, an ineffective attempt at beautifying and obfuscating something whose unsightly features are still evident. We need a fundamental, structural, and thoughtful shift in our paradigm beginning with hypotheses driven by meaningful questions and careful selection of informative characters for investigation. We need a return—or rather, beginning—to investigating real human biological variation.

Author Contributions: A.H.R., S.E.W. conceptualized the paper, collected the data, and contributed to writing and editing the manuscript. A.H.R. performed the statistical analyses. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data a contained within the article.

Acknowledgments: We thank Kate Spradley for providing the Guatemalan sample, and John Fredy Ramirez for access to the Antioquia Modern Skeletal Reference Collection in Medellín, Colombia. We thank Elizabeth DiGangi and Jonathan Bethard for making the Colombian data available and for comments on the manuscript. We also thank Chris Kligenberg for MorphoJ guidance.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Langley, N.R.; Tersigni-Tarrant, M.A. *Forensic Anthropology: A Comprehensive Introduction*; CRC Press: Boca Raton, FL, USA, 2017.
2. Smay, D.; Armelagos, G. Galileo Wept: A Critical Assessment of the Use of Race in Forensic Anthropology. *Transform. Anthropol.* **2000**, *9*, 19–29. [CrossRef]
3. Bethard, J.D.; DiGangi, E.A. Letter to the Editor—Moving Beyond a Lost Cause: Forensic Anthropology and Ancestry Estimates in the United States. *J. Forensic Sci.* **2020**, *65*, 1791–1792. [CrossRef] [PubMed]
4. DiGangi, E.A.; Bethard, J.D. Uncloaking a Lost Cause: Decolonizing ancestry estimation in the United States. *Am. J. Phys. Anthropol.* **2021**, *175*, 422–436. [CrossRef]
5. Edgar, H.J.; Hunley, K.L. Race reconciled?: How biological anthropologists view human variation. *Am. J. Phys. Anthropol.* **2009**, *139*, 1–4. [CrossRef] [PubMed]
6. Lieberman, L.; Stevenson, B.W.; Reynolds, L.T. Race and Anthropology: A Core Concept without Consensus. *Anthropol. Educ. Q.* **1989**, *20*, 67–73. [CrossRef]
7. Sauer, N.J. Forensic Anthropology and the Concept of Race—If Races Don't Exist, Why Are Forensic Anthropologists So Good at Identifying Them. *Soc. Sci. Med.* **1992**, *34*, 107–111. [CrossRef]
8. Ousley, S.; Jantz, R.; Freid, D. Understanding race and human variation: Why forensic anthropologists are good at identfying race. *Am. J. Phys. Anthropol.* **2009**, *139*, 68–76. [CrossRef]
9. Stull, K.E.; Bartelink, E.J.; Klales, A.R.; Berg, G.E.; Kenyhercz, M.W.; L'Abbé, E.N.; Mariscal, C. Commentary on: Bethard JD, DiGangi EA. Letter to the Editor-Moving beyond a lost cause: Forensic anthropology and ancestry estimates in theUnited States. *J. Forensic Sci.* **2021**, *66*, 417–420. [CrossRef] [PubMed]
10. Goodman, A.H. Bred in the Bone? *Sciences* **1997**, *37*, 20–25. [CrossRef]
11. Blakey, M.L. Scientific racism and the biological concept of race. *Lit. Psychol.* **1999**, *45*, 29.
12. Albanese, J.; Saunders, S.R. Is it possible to escape racial typology in forensic identification? In *Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death*; Schmitt, A., Cunha, E., Pinheiro, J., Eds.; Humana: Ttowa, NJ, USA, 2006; pp. 281–316.
13. Maier, C.; Craig, A.; Adams, D.M. Language use in ancestry research and estimation. *J. Forensic Sci.* **2021**, *66*, 11–24. [CrossRef]
14. Ross, A.H.; Slice, D.E.; Ubelaker, D.H.; Falsetti, A.B. Population affinities of 19th Century Cuban crania: Implications for identification criteria in South Florida Cuban Americans. *J. Forensic Sci.* **2002**, *49*, 1–6. [CrossRef]
15. Brues, A.M. Forensic diagnosis of race—General race vs specific populations. *Soc. Sci. Med.* **1992**, *34*, 125–128. [CrossRef]
16. Caspari, R. Race, then and now: 1918 revisited. *Am. J. Phys. Anthropol.* **2018**, *165*, 924–938. [CrossRef]
17. Von Cramon-Taubadel, N. Evolutionary insights into global patterns of human cranial diversity: Population history, climatic and dietary effects. *J. Anthropol. Sci.* **2014**, *92*, 43–77. [PubMed]
18. Sneath, P.H.; Sokal, R.R. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*; Stony Brook University: Stony Brook, NY, USA, 1975.
19. Ross, A.H. Regional isolation in the Balkan region: An analysis of craniofacial variation. *Am. J. Phys. Anthropol.* **2004**, *124*, 73–80. [CrossRef]
20. Coon, C.S. *The Origin of Races*; Alfred A. Knopf.: New York, NY, USA, 1962.
21. Gill, G.W.; Rhine, S. *Skeletal Attribution of Race: Methods for Forensic Anthropology*; Maxwell Museum of Anthropology: Albuquerque, NM, USA, 1990.
22. Sardi, M.L.; Rozzi, F.R.; Dahinten, S.L.; Pucciarelli, H.M. Amerindians: Testing the hypothesis about their homogeneity. *Comptes Rendus Palevol* **2004**, *3*, 403–409. [CrossRef]
23. Cardini, A.; O'Higgins, P.; Rohlf, F.J. Seeing Distinct Groups Where There are None: Spurious Patterns from Between-Group PCA. *Evol. Biol.* **2019**, *46*, 303–316. [CrossRef]

24. Rolf, F.J. Why Clusters and Other Patterns Can Seem to be Found in Analyses of High-Dimensional Data. *Evol. Biol.* **2021**, *48*, 1–16. [CrossRef]
25. Ross, A.H.; Williams, S.E. Testin Repeatability and Error of Coordinate Landmark Data Acquired from Crania. *J. Forensic Sci.* **2008**, *53*, 782–785. [CrossRef]
26. Stephen, A.J.; Wegscheider, P.K.; Nelson, A.J.; Dickey, J.P. Quantifying the precision and accuracy of the MicroScribe G2X three-dimensional digitizer. *Digit. Appl. Archaeol. Cult. Herit.* **2015**, *2*, 28–33. [CrossRef]
27. Slice, D.E.; Ross, A.H. *3D-ID: Geometric Morphometric Classification of Crania for Forensic Scientists*; NCJRS: Washington, DC, USA, 2009.
28. Slice, D.E. Morpheus et al. Java Edition. Available online: http:/morphlab.sc.fsu.edu/ (accessed on 28 June 2021).
29. Klingenberg, C.P. *MorphoJ*: An integrated software package for geometric morphometrics. *Mol. Ecol. Resour.* **2011**, *11*, 353–357. [CrossRef] [PubMed]
30. Kendall, D.G. The diffusion of shape. *Adv. Appl. Prob.* **1977**, *9*, 428–430. [CrossRef]
31. Zelditch, M.L.; Swiderski, D.L.; Sheets, H.D. *Geometric Morphometrics for Biologists: A Primer*; Academic Press: Salt Lake City, UT, USA, 2012.
32. *JMP*, Version 14.1 ed; SAS Institute Inc.: Cary, NC, USA, 1989.
33. Sokal, R.R. The continental population structure of Europe. *Annu. Rev. Anthropol.* **1991**, *20*, 119–140. [CrossRef]
34. Pietrusewsky, M. Traditional morphometrics and biological distance methods and an example. In *Biological Anthropology of the Human Skeleton*, 3rd ed.; Katzenberg, M.A., Grauer, A.L., Eds.; Wiley-Blackwell: Hoboken, NJ, USA, 2019; pp. 547–591.
35. Sokal, R.R.; Uytterschaut, H. Cranial variation in European populations: A spatial autocorrelation study at three time periods. *Am. J. Phys. Anthropol.* **1987**, *74*, 21–38. [CrossRef]
36. Anselin, L.; Syabri, I.; Kho, Y. GeoDa: An Introduction to Spatial Data Analysis. In *Handbook of Applied SpatialAanalysis*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 73–89.
37. Hefner, J.T.; Ousley, S.D. Statistical Classification Methods for Estimating Ancestry Using Morphoscopic Traits. *J. Forensic Sci.* **2014**, *59*, 883–890. [CrossRef] [PubMed]
38. Jantz, R.L.; Ousley, S.D. *Fordisc 3.1: Computerized Forensic Discriminant Functions*; The University of Tennessee: Knoxville, TN, USA, 2005.
39. Navega, D.; Coelho, C.; Vicente, R.; Ferreira, M.T.; Wasterlain, S.; Cunha, E. AncesTrees: Ancestry estimation with randomized decision trees. *Int. J. Leg. Med.* **2015**, *129*, 1145–1153. [CrossRef]
40. Ross, A.H.; Ubelaker, D.H. Complex Nature of Hominin Dispersals: Ecogeographical and Climatic Evidence for Pre-Contact Craniofacial Variation. *Sci. Rep.* **2019**, *9*, 1–9. [CrossRef] [PubMed]
41. Suárez, O.J. *La población del Istmo de Panamá: Estudio de geohistoria*; Ediciones de Cultura Hispánica: Madrid, Spain, 1998.
42. Salzano, F.M.; Bortolini, M.C. *The Evolution and Genetics of Latin American Populations*; Cambridge University Press: Cambridge, UK, 2005.
43. Price, T.D.; Tiesler, V.; Zabala, P.; Coppa, A.; Freiwald, C.; Schroeder, H.; Cucina, A. (Ahead of Print). Home Is the Sailor: Investigating the Origins of the Inhabitants of La Isabela, the First European Settlement in the New World. *Curr. Anthropol.* **2020**, *61*, 5. [CrossRef]
44. Ross, A.H.; Keegan, W.F.; Pateman, M.; Young, C.B. Faces Divulge the Origins of Caribbean Prehistoric Inhabitants. *Sci. Rep.* **2020**, *10*, 147–149. [CrossRef]
45. Panama Canal Commission. *The Panama Canal Review*; October 1 Issue; 1979. Available online: https://ufdcimages.uflib.ufl.edu/UF/00/09/73/66/00059/UF00097366_00059.pdf (accessed on 28 June 2021).
46. Chiappetta, R.F.; Treleaven, T.; Nixon, E.; Smith, J.D. History and expansion of the Panama Canal. *Fragblast* **1998**, *2*, 313–340. [CrossRef]
47. McCullough, D. *The Path Between the Seas: The Creation of the Panama Canal*; Simon & Schuster: New York, NY, USA, 1977; pp. 1870–1914.

biology

MDPI

Article

Bone Diagenesis in Short Timescales: Insights from an Exploratory Proteomic Analysis

Noemi Procopio [1,*,†], Caley A. Mein [2,3,†], Sefora Starace [4], Andrea Bonicelli [1] and Anna Williams [2,3]

1 Forensic Science Research Group, Faculty of Health and Life Sciences, Applied Sciences, Northumbria University, Newcastle Upon Tyne NE1 8ST, UK; andrea.bonicelli@northumbria.ac.uk
2 School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK; CAMein@uclan.ac.uk (C.A.M.); AWilliams34@uclan.ac.uk (A.W.)
3 School of Natural Sciences, University of Central Lancashire, Preston PR1 2HE, UK
4 Dipartimento di Chimica, University of Turin, Via P. Giuria 7, 10125 Turin, Italy; sefora.starace@edu.unito.it
* Correspondence: noemi.procopio@northumbria.ac.uk
† Authors contributed equally to first co-authorship.

Simple Summary: Understanding the origin of bone degradation led by bacterial decomposition is essential in order to allow for the creation of better models to estimate the time elapsed since death for forensic casework, as well as for the preservation of archaeological specimens over the course of time. Within this study we applied modern proteomic technologies in order to evaluate how proteins in decomposing rat bones are affected by different post-mortem conditions, such as different depositional environments (buried versus exposed samples) and different sample types (whole carcasses versus fleshed limbs versus defleshed bones), over a period of 28 weeks. We found that the abundance of specific proteins was associated either with a microbial-led type of decomposition or with a specific post-depositional environment. Overall, this study shows that proteomic analyses can be useful to identify microbially- versus non-microbially driven decomposition, and that specific proteins—such as bone marrow and plasma proteins—can be more affected than others by extrinsic agents, whereas calcium-binding proteins seem to be more affected by microbial degradation.

Abstract: The evaluation of bone diagenetic phenomena in archaeological timescales has a long history; however, little is known about the origins of the microbes driving bone diagenesis, nor about the extent of bone diagenesis in short timeframes—such as in forensic contexts. Previously, the analysis of non-collagenous proteins (NCPs) through bottom-up proteomics revealed the presence of potential biomarkers useful in estimating the post-mortem interval (PMI). However, there is still a great need for enhancing the understanding of the diagenetic processes taking place in forensic timeframes, and to clarify whether proteomic analyses can help to develop better models for estimating PMI reliably. To address these knowledge gaps, we designed an experiment based on whole rat carcasses, defleshed long rat bones, and excised but still-fleshed rat limbs, which were either buried in soil or exposed on a clean plastic surface, left to decompose for 28 weeks, and retrieved at different time intervals. This study aimed to assess differences in bone protein relative abundances for the various deposition modalities and intervals. We further evaluated the effects that extrinsic factors, autolysis, and gut and soil bacteria had on bone diagenesis via bottom-up proteomics. Results showed six proteins whose abundance was significantly different between samples subjected to either microbial decomposition (gut or soil bacteria) or to environmental factors. In particular, muscle- and calcium-binding proteins were found to be more prone to degradation by bacterial attack, whereas plasma and bone marrow proteins were more susceptible to exposure to extrinsic agents. Our results suggest that both gut and soil bacteria play key roles in bone diagenesis and protein decay in relatively short timescales, and that bone proteomics is a proficient resource with which to identify microbially-driven versus extrinsically-driven diagenesis.

Keywords: taphonomy; bone proteomics; PMI; microbial decomposition; bioerosion; forensic sciences

Citation: Procopio, N.; Mein, C.A.; Starace, S.; Bonicelli, A.; Williams, A. Bone Diagenesis in Short Timescales: Insights from an Exploratory Proteomic Analysis. *Biology* **2021**, *10*, 460. https://doi.org/10.3390/biology10060460

Academic Editors: Ann H. Ross and Eugénia Cunha

Received: 20 April 2021
Accepted: 20 May 2021
Published: 23 May 2021

1. Introduction

Bones are among the longest-preserved biological tissues in nature. Nevertheless, it is known that their preservation in both exposed conditions and buried contexts is affected by a multitude of extrinsic environmental variables, including physical and chemical environmental agents, scavengers, soil hydrology and pH, temperature, and microorganism-driven bioerosion [1]. Bone diagenetic processes have been extensively investigated over long timescales (e.g., in the archaeological records), with the aim of better understanding the taphonomic processes driving bones' preservation and their ultimate conversion into fossil specimens [2–4], as well as the origin of the microbes driving bioerosion [5]. In particular, the effects of microbial attack and bone hydrolysis, resulting from intrinsic (i.e., derived from gut) or extrinsic (i.e., derived from soil and environment) microorganisms on the bone structure, are still a debated topic in archaeology and in palaeontology, and researchers have not yet fully agreed on which source can be considered to be the major driver for bone bioerosion [1,6]. In addition, these processes are less well understood when considering shorter timeframes (e.g., in forensic contexts), even though early taphonomy studies may benefit a wide range of fields, including forensic sciences, in addition to archaeology, biomolecular archaeology, and palaeontology.

Biomolecules in bones have been successfully found in a variety of archaeological and fossil records [7–10], as well as in forensic contexts (e.g., human bones collected from caseworks or cemeteries) [11,12], or in situations simulating forensic scenarios (e.g., using animal proxies to conduct bone proteomic studies) [13,14]. In particular, proteins have been shown to survive better than DNA [9,15]. Bone proteomics is a promising tool with which to evaluate chronological bone degradation through the evaluation of the survival of collagen [16] and of the decay of non-collagenous proteins (NCPs) [17–19]. Several studies show that bone collagen content can be halved in less than a thousand years [20], depending on burial conditions, with microbial attack having a strong impact on its persistence, particularly for bones that have been buried/located in relatively cold regions, such as northern Europe [21]. In neutral conditions, bone collagen is considered to be predominantly stable and insoluble [22]. On the one hand, collagen is protected by the strong linkage with the mineral matrix of the bone, and on the other hand, it protects the hydroxyapatite from dissolution [22]. Collagen deterioration through hydrolysis is a complex process mainly driven by bacterial collagenases, despite the fact that chemical (non-enzymatical) hydrolysis can also happen in situations involving the presence of high temperatures or extreme pHs [23]. Enzymatically-driven hydrolysis requires the access of collagenases to the collagen helix—a situation that occurs only when there is enough space among crystallites to allow the penetration of the enzymes [23], such as when the matrix has been partially dissolved by the presence of, for example, acid metabolites generated by the putrefaction of the soft tissues during the cadaveric decomposition [22]. Bacterial proteases are not able to function at acidic pHs, so the environment should become more alkaline/neutral to allow collagenases in bones to work [20].

Despite the great number of works aimed at understanding the mechanisms behind collagen survival within bones, less is known regarding the degradation and subsistence of NCPs [24]. Various studies on archaeological and palaeontological bones suggested that NCPs may survive better than collagen, due to their strong affinity to the mineral matrix of the bone [17,25,26]. However, there are limited experimental studies exploring the mechanisms and the quantitative degradation of NCPs in forensic timeframes. A work based on the burial of swine carcasses showed that the majority of changes in the proteome occur in the first four months of the PMI [13]. The study highlighted two different behaviours: proteins for which the relative abundance significantly decreased during the first months, to then become more stable (plasma proteins); and proteins showing some degree of stability in the initial stages of decomposition, followed by a drastic decrement (muscle proteins) [13]. These findings are also supported by a study by Creamer and Buck [27], who employed luminol to prove the time-dependent degradation of haemoglobin—a methodology that can be applied to classify skeletal human remains as forensic or archaeological.

In addition to haemoglobin, transferrin has always attracted attention, as it has been considered a reliable indicator of PMI, although it is not directly extracted from bone tissue [28]. Despite its potential, its degradation was seen to be highly temperature-dependent [28] and, therefore, not the most suitable biomarker for long PMIs when temperature can fluctuate considerably. The degradation of the muscle proteins identified in bone samples was previously interpreted as a link to the complete decomposition of superficial soft tissues at prolonged PMIs [13]. This finding is also supported in the literature, where similar patterns were seen for muscle proteins extracted from muscular tissue after various PMIs [29,30].

In order to understand the phenomena involved in the degradation of proteins in bones, it is essential to illustrate the changes to which a corpse is subjected from its death, as well as the role of microbial communities—also referred to by certain authors as the necrobiome—which are among the main agents involved in post-mortem tissue decomposition [31]. The first late post-mortem change is the autolytic process, which takes place just after death and is caused by the breakdown of cellular membranes, which causes the release of hydrolytic enzymes able to digest the surrounding tissues. During this process, the body environment is rapidly converted from an aerobic to an anaerobic one, allowing anoxic bacteria from the gut (endogenous bacteria) to multiply and the body to enter the putrefactive stage, which generally starts 1 hour post-mortem and lasts for 48 h [6]. Throughout putrefaction, the second stage of late post-mortem changes, bacteria transmigrate from the gut to the rest of the body via the blood vessels in a matter of hours [32,33], reaching the bones within a day of death [22]. The activity of endogenous bacteria via reductive catalysis results in bloating of the corpse, due to the accumulation of gases [22] that would eventually lead to abdominal rupture and the exposure of internal tissues, offering an ideal environment for exogenous microbial communities present in the soil and in the surrounding air [34–36]. These exogenous communities proliferate due to the richness of the nutrients offered by the decomposing corpse, replacing the endogenous communities [37]. With the progression of decomposition, specific bacterial species are attracted in a quite predictable way, regardless of the type of burial environment. For this reason, the microbial succession in soil has been studied by several authors as a means of estimating the PMI [38,39].

Although the post-mortem microbial succession has attracted significant interest for PMI estimation, the origin of the bacteria responsible for bone diagenesis is still hugely debated [1,5,6]. A study conducted by Damann et al. [40] identified changes in the abundances of Firmicutes, Actinobacteria, Acidobacteria, and Proteobacteria within human ribs at various decomposition stages. In particular, they found a predominance of gut bacteria during the earlier decomposition stages (e.g., partially skeletonised remains), and of soil bacteria during the advanced decomposition stages (e.g., dry remains) [40]. There is a consensus in the current literature that microbes (both bacteria and fungi) can affect the internal microstructure of bones [1,41,42]. White and Booth [1] proposed gut bacteria as the main drivers of the bone diagenetic processes; Reiche et al. [43] suggested instead that soil microbes are mainly responsible for bone diagenesis in archaeological samples. However, given the depth of burials, and the reduced number of soil microbial communities found at the average burial depth in contrast with shallower burial depths [44], it may be possible that soil microbes alone are not uniquely responsible for the degree of diagenesis observed in the work of Reiche et al. [43]. Jans [22] seems to attribute more importance to endogenous bacteria as drivers for bone bioerosion, whereas, more recently, Turner-Walker [5] suggested that soil bacteria are essential for bone diagenesis, and stated that the "hypothesis that bacterial tunnelling arises from gut bacteria, although plausible, is unproven".

One aspect that should never be neglected is the importance of the depositional environment, which can also lead to measurable physical and chemical changes to the microstructure of the bone, due to microbial interactions between the hard tissue and the surrounding environment [43,45]. This introduces a large number of factors that have to be considered when evaluating the decomposition and decay of a corpse; therefore,

systematic experiments are necessary in order to limit these variables and isolate the ones of interest. Thus, ultimately, both the decomposition process [46,47] and the depositional environment [37,47,48] can have strong effects on the extent of bone diagenesis, which seems to occur at different rates in buried bones when compared with exposed ones [49].

The present work represents a preliminary study that aims to fill the knowledge gaps concerning early bone diagenesis in forensic timeframes through the innovative approach of bone proteomics. We consider here the existence of three main microbial groups: endogenous microbes (derived predominantly from the gut), aerobic environmental microbes (present in the air surrounding exposed carcasses), and soil microbes (exogenous bacteria in the burial environment). We focused specifically on the proteomic alterations observed in bones exposed predominantly to the combined action of autolysis coupled with the effects that endogenous and environmental bacteria, soil bacteria, or environmental bacteria alone can have on bodies, over short timescales. It should be noted that the presence of the "environmental bacteria" also implies the contemporaneous presence of other living species (e.g., insects) able to colonise the decomposing bodies. The deposition times ranged between 4 and 28 weeks, and the combination between different "sample types" (whole remains, excised fleshed limbs, and defleshed bones) and "depositional environments" (exposed on a clear plastic surface in an outdoor environment or buried in soil) allowed the evaluation of the effects that these different microbial groups had on the bone proteomes. The objective of this approach is to evaluate the presence of proteomic alterations over short/forensic timescales, and assess the effects that specific decomposers and environmental variables have on them. We were able to clarify the potential that bone proteomics has in the evaluation of early diagenetic processes, and to elucidate some of the mechanisms underlying the NCPs' preservation within bones in specific conditions within short timescales.

2. Materials and Methods

2.1. Field Experiment and Sample Composition

Eleven black rats (*Rattus Rattus*) were purchased frozen from a reptile centre (Northampton Reptile Centre), operating in accordance with the Animal Welfare Act 2006, compliant with ethical research standard practices. We used animal models in order to test different post-mortem conditions with minimal interindividual variability. The animals were flash frozen within ~2 h after death, delivered frozen, and stored at −20 °C until the beginning of the study. Five rats (named "whole body" and "control" for the purpose of this paper) were not subjected to any dissection, so they were intact until the beginning of the experiment. Six rats were used to obtain either back limbs (named "fleshed limb") or back limb bones (named "defleshed bone"). Prior to dissection, they were defrosted overnight at 4 °C. Although we are aware of the effects that freeze-thawing can have on intrinsic bacterial communities [50] and on soft tissues and cell structures [51], several studies have showed that bacteria are able to survive during freezing procedures [52,53], so we are confident that this experimental choice has not impaired the correct evaluation of the decomposition phenomena associated with the presence of intrinsic bacteria in this study.

The experimental samples were buried at the University of Huddersfield's animal taphonomic facility (HuddersFIELD), from mid-November 2018 to May 2019. The facility is situated on grassy farmland in West Yorkshire (UK). Local temperature and rainfall information was collected using a local weather station, World Weather Online, and can be found at (https://www.worldweatheronline.com/halifax-weatheraverages/west-yorkshire/gb.aspx, accessed on 1 November 2019). Average monthly temperature and rainfall readings/values were recorded from November 2018 to May 2019—the duration of the field experiments (Supplementary Materials, Figure S1).

The samples were deposited in either exposed or buried conditions. The exposed samples were placed into two large plastic boxes, one box for each tissue type (e.g., one box for the whole rats, one for the excised fleshed limbs). The boxes were open, had holes

drilled in the base to allow drainage of rainwater, and were placed within large, locked wire cages in order to protect them from large scavengers. The buried samples were placed into smaller, soil-filled 3-L boxes, in order to prevent disturbance of the samples by scavengers. The samples were placed on top of ~1.35 L of Godwins topsoil and further covered with 1.35 L of the same topsoil. These boxes were closed and their lids were weighted down to protect the contents from scavengers. Upon collection, the defleshed bones were cleaned of any soil or debris using distilled water. The whole remains and the excised fleshed limbs were defleshed where necessary using a #10 scalpel. All bones were cleaned using distilled water and then stored in the freezer until further preparation. Each bone specimen was placed into a small clean polythene bag and gently broken into smaller fragments with a mortar. Three small fragments (sized ~3–4 mm^2) were used as biological replicates (named "A", "B", and "C") for each of the samples used (Table 1). As a result, we totalled 39 samples that were subjected to further proteomic extraction and analysis.

Table 1. Samples used for proteomics, with the sample codes, depositional environment, sample type, post-mortem timescale, and proteomics code associated with them.

Sample Codes	Depositional Environment	Sample Type	Timescale in Weeks	Proteomics Code
CTRL A-B-C	Control	Control	0	NP25-26-27
w12 BB A-B-C	Buried	Defleshed	12	NP28-29-30R
w12 EW A-B-C	Exposed	Whole body	12	NP31-32-33
w12 EF A-B-C	Exposed	Fleshed limb	12	NP19-20-21
w20 BB A-B-C	Buried	Defleshed	20	NP4-5-6
w20 EW A-B-C	Exposed	Whole body	20	NP1-2-3
w20 EF A-B-C	Exposed	Fleshed limb	20	NP7-8-9
w24 BB A-B-C	Buried	Defleshed	24	NP34-35-36
w24 EW A-B-C	Exposed	Whole body	24	NP37-38-39
w24 EF A-B-C	Exposed	Fleshed limb	24	NP22-23-24
w28 BB A-B-C	Buried	Defleshed	28	NP10-11-12
w28 EW A-B-C	Exposed	Whole body	28	NP13-14-15
w28 EF A-B-C	Exposed	Fleshed limb	28	NP16-17-18

2.2. Protein Extraction

Proteins were extracted following the protocol proposed by Procopio and Buckley [54]. Briefly, each bone fragment was placed in a separated 1.5-mL tube and decalcified with 1 mL of 10% formic acid (Fisher Scientific, Loughborough, UK) for 6 hours at 4 °C. The acid-soluble fraction was discarded, and the pellet was then treated with 6 M guanidine hydrochloride (Sigma-Aldrich, Gillingham, UK)/100 mM TRIS buffer (Thermo Fisher Scientific, Paisley, UK) with an adjusted pH of 7.4 for 18 h at 4 °C. The acid-insoluble fraction was then exchanged with 50 mM ammonium acetate (Scientific Laboratories Supplies, Nottingham, UK) using 10 kDa molecular weight cut-off filters (Vivaspin®500 from Sartorius, Göttingen, Germany), and the proteins were then reduced with 5 mM dithiothreitol (Fluorochem, Hadfield, UK) for 40 min at room temperature, alkylated using 15 mM iodoacetamide (Sigma-Aldrich, UK) for 45 min in the dark at room temperature, and quenched by adding 5 mM dithiothreitol. Proteins were digested using trypsin (Promega, Southampton, UK) for 5 hours at 37 °C, and then frozen. We added 1 v/v% trifluoroacetic acid (Fluorochem, Hadfield, UK) to the samples prior to their desalting and purification, in order to bring them to a final concentration of 0.1 v/v% trifluoroacetic acid. The purification was achieved using OMIX C18 pipette tips (Agilent Technologies, Stockport, UK), in accordance with the manufacturer's instructions. The peptides were finally eluted into 100 µL of 50 v/v% acetonitrile (Thermo Fisher Scientific, UK)/0.1 v/v% trifluoroacetic acid. Samples were then left in the fume cupboard with lids open in order to dry them completely at room temperature prior to their submission for LC–MS/MS analysis.

2.3. LC–MS/MS Analysis

Samples resuspended in 5 v/v% ACN/0.1 v/v% TFA were analysed via LC–MS/MS using an Ultimate™ 3000 Rapid Separation LC (RSLC) nano LC system (Dionex Corporation, Sunnyvale, CA, USA) coupled to a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Peptides were separated on an EASY-Spray™ reverse phase LC Column (500 mm × 75 µm diameter (i.d.), 2 µm, Thermo Fisher Scientific, Waltham, MA, USA) using a gradient from 96 v/v% A (0.1 v/v% FA in 5 v/v% ACN) and 4 v/v% B (0.1 v/v% FA in 95 v/v% ACN) to 8 v/v%, 30 v/v%, and 50% B at 14, 50, and 60 min, respectively, at a flow rate of 300 nL min^{-1}. Acclaim™ PepMap™ 100 C18 LC Column (5 mm × 0.3 mm i.d., 5 µm, 100 Å, Thermo Fisher Scientific) was used as trap column at a flow rate of 25 µL min-1 kept at 45 °C. The LC separation was followed by a cleaning cycle with an additional 15 min of column equilibration time; then, peptide ions were analysed in full-scan MS scanning mode at 35,000 MS resolution with an automatic gain control (AGC) of 1e6, injection time of 200 ms, and scan range of 375–1400 m/z. The top 10 most abundant ions were selected for data-dependent MS/MS analysis with a normalized collision energy (NCE) level of 30, performed at 17,500 MS resolution with an AGC of 1×10^5 and maximum injection time of 100 ms. The isolation window was set to 2.0 m/z, with an underfilled ratio of 0.4%, and dynamic exclusion was employed; thus, one repeat scan (i.e., two MS/MS scans in total) was acquired in a 45 s repeat duration, with the precursor being excluded for the subsequent 45 s.

2.4. Protein Identification and Statistical Analysis

Peptide mass spectra were then searched against the SWISS-PROT database using the Mascot search engine (version 2.5.1; www.matrixscience.com, accessed on 10 December 2019) for matches to primary protein sequences, specifying the taxonomy filter as *Rattus rattus*. This search included the fixed carbamidomethyl modification of cysteine, which results from the addition of DTT to proteins. In the light of our aims, deamidation (asparagine and glutamine) and oxidation (lysine, methionine and proline) were also considered as variable modifications. The enzyme was set to trypsin, with a maximum of two missed cleavages allowed. It was assumed that all spectra held either 2+ or 3+ charged precursors.

Progenesis QI for Proteomics software (version 4.2; Nonlinear Dynamics, Newcastle, UK) was used to identify the proteins and assess their relative abundance in each sample. The relative abundance of the proteins present in the samples is calculated by the software by measuring the peptide ion abundances as a result of the sum of the areas under the curve (AUC) for each peptide ion. The software normalizes each LC–MS/MS run against a reference run automatically selected as the normalization reference, in order to consider and to correct the systematic experimental variations that can occur between different runs; then, it enables protein comparisons between different experimental conditions, and the identification of protein expression changes. Peptide ions with a score of <23, indicating identity or extensive homology ($p < 0.05$), were excluded from the analysis based on the Mascot evaluation of the peptide score distribution. To further increase the reliability of the obtained results, proteins with a peptide count of <2 were also excluded from further analysis.

STRING software (version 11.0) was used to visualize links among the identified proteins, and to evaluate the significance of their interactions [55]. The confidence score set for showing interactions was set to "medium = 0.400".

Statistical analysis was carried out on arcsinh normalised data [56], in the same way in which the Progenesis software operates. Principal component analysis (PCA) was performed on the normalized abundance values exported from Progenesis, using R software with the factoextra and ggplot2 packages, using proteins sorted by their ANOVA FDR adjusted p values (or "q value") in order to exclude the proteins that showed similar relative abundances across different conditions (and that did not contribute to cluster separations). Protein abundances were considered to have significantly contributed to the explanation of the variance between different groups of samples when their ANOVA

$p < 0.05$. For post-hoc analysis, Tukey's HSD test was employed with significance set at $p < 0.05$ for pairwise comparison.

3. Results

Overall, 67,490 MS/MS spectra were acquired from the LC–MS/MS analyses, and were searched against the SWISS-PROT database using Mascot. After the refinement steps previously mentioned, 2,539 search results (ions) were matched, and after the exclusion of proteins matched with less than 2 unique peptides, we obtained 168 proteins. In order to exclude the possibility that the proteomic results were associated with differences in the PMIs of the specimens, we grouped samples exclusively based on their PMIs, and plotted them on a PCA graph (Figure 1A). The samples did not cluster in clearly and defined positions, apart from the control samples, and the variability explained by summing the first and second components was less than 50%. To verify that the clustering was not affected by the inclusion of the control group in the model, we excluded those from the PCA (Figure 1B). Moreover, in this case, no clear clusters were identified, and samples belonging to different PMIs overlapped on both dimensions without any clear trend.

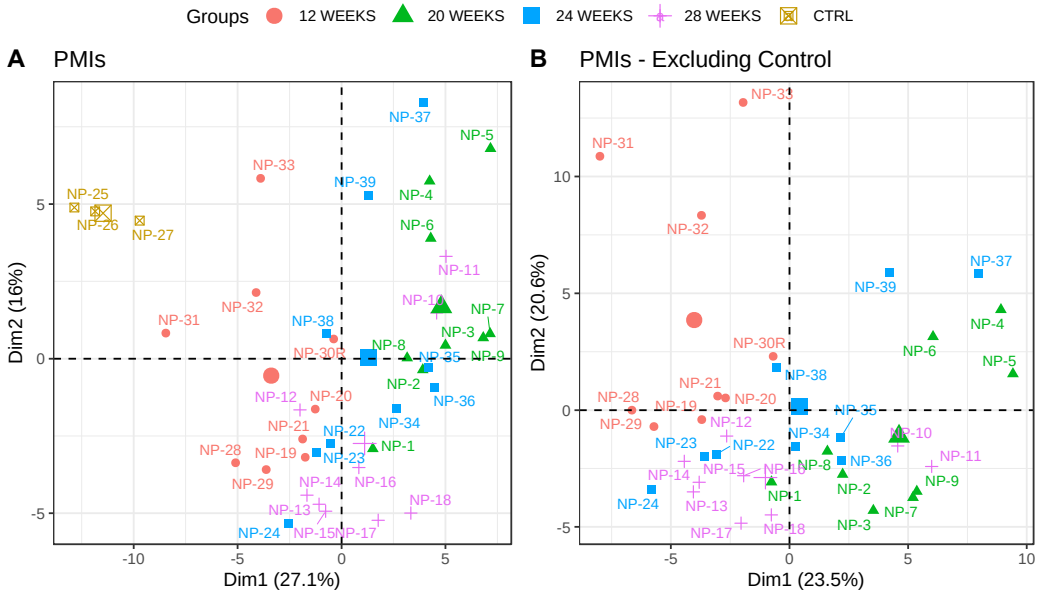


Figure 1. PCA plot representing the samples grouped by their PMIs; (**A**) including control samples, and (**B**) excluding control samples. The same-coloured icons indicate the same PMI, as reported in the legend. Control (CTRL) = 0 weeks PMI. Protein abundances used for the PCAs were the ones whose q values were significant between the different conditions ($n = 90$, $q < 0.05$). The bigger icons in the plot for each group represent the centroids of the samples in the group.

Samples were then grouped based on the depositional environment and on the sample type, in order to test whether or not the bone proteome was affected by it and, consequently, whether any variation in bone diagenesis could be observed via the analysis of the bone proteome. Results showed a significant difference in the proteomes of the three distinct sample types, and it was possible to identify specific proteins responsible for the variance observed, particularly when comparing the exposed fleshed samples with either the buried defleshed bones or the exposed whole bodies (Figure 2). The only situation in which it was not possible to identify proteins with significant q values was the exposed whole bodies versus the buried defleshed bones, meaning that the two groups shared a similar proteomic profile. The reported PCAs were able to cluster both fleshed and defleshed limbs (Figure 2B), and fleshed limbs versus whole bodies (Figure 2C), with models able to explain 58.4% and 61.2%, respectively, of the total variance between the two groups. We were also able to find the specific proteins that contributed the most to the variance observed in the PCAs (Figure 2E,F; red arrows).

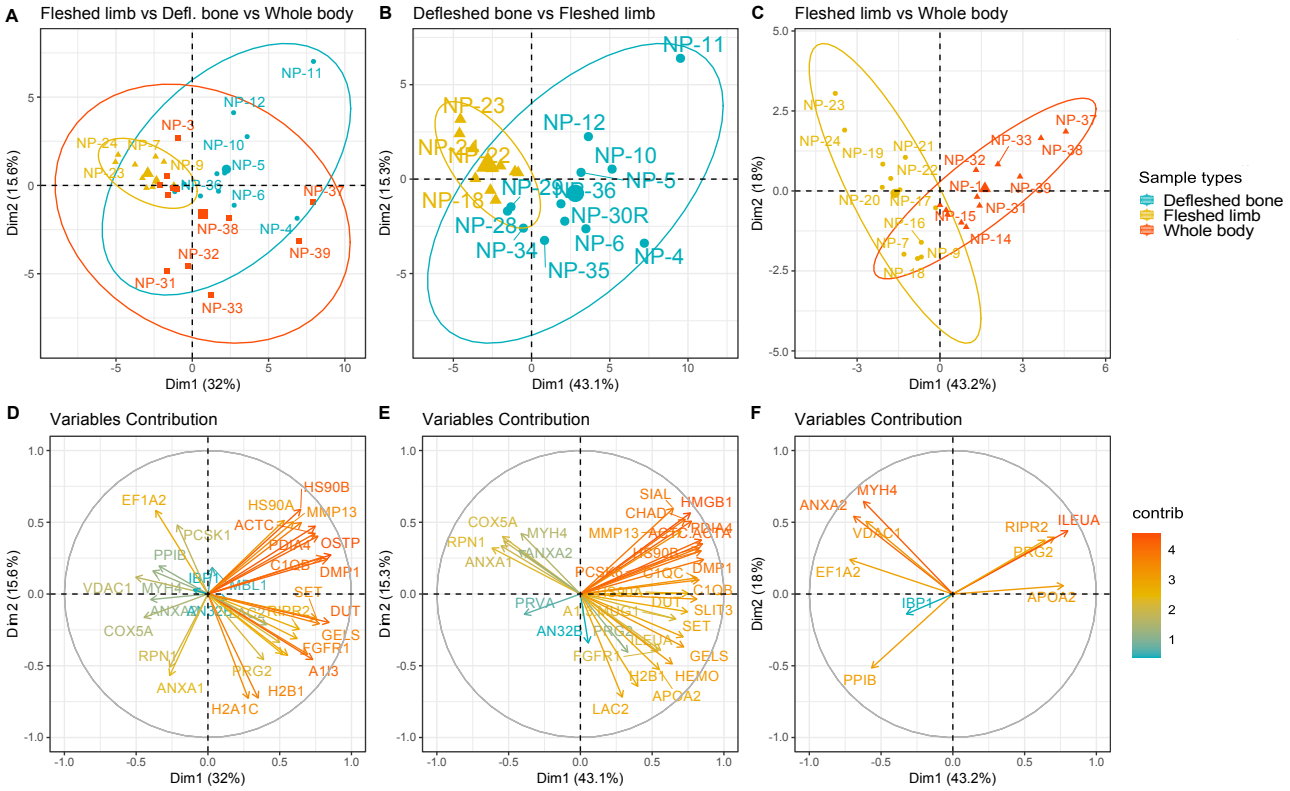


Figure 2. PCA plots (top) and variable maps (bottom) representing the samples grouped respectively in: (**A**,**D**) buried defleshed bones vs. exposed fleshed limbs; (**B**,**E**) exposed fleshed limbs vs. exposed whole bodies; and (**C**,**F**) buried defleshed bones vs. exposed fleshed limbs vs. exposed whole bodies. In each cluster, a bigger coordinate with the same colour was calculated as the centroids of the samples in the group. The same-coloured icons indicate the same condition. The ellipse shows a cluster categorised by the same condition (confidence interval of 0.95). Protein abundances used for the PCAs were the ones that were statistically significant according to their q values.

Starting from the significant contributors that were able to discriminate different groups of samples, we selected the ones that were significant ($q < 0.05$) both for fleshed limbs versus defleshed bones and for fleshed limbs versus whole bodies. We obtained six proteins in this manner, which differed in abundance when tested by post-hoc pairwise comparisons: apolipoprotein A-II (APOA2), leukocyte elastase inhibitor A (ILEUA), bone marrow proteoglycan (PRG2), annexin A2 (ANXA2), voltage-dependent anion-selective channel protein 1 (VDAC1), and myosin-4 (MYH4) (Figure 3). APOA2 is a plasma protein produced by the liver and commonly found in bone tissue [57]; ILEUA and PRG2 are bone marrow proteins [58,59]; ANXA2 is a bone specific protein associated with osteoclasts and bone tissue formation [60]; and MYH4 and VDAC1 are both skeletal muscle proteins [61,62]. When uploaded on the STRING software, these six proteins did not show significantly more interactions then expected, as predictable from the very small set of proteins ($n = 6$). However, according to Gene Ontology (GO) terms, we found functional enrichments for the molecular functions "binding" ($n = 5$; strength 0.56), "protein-containing complex binding" ($n = 4$; strength 1.45), and "anion binding" ($n = 4$; strength 1.04), for the biological processes "regulation of cellular metabolic process" ($n = 5$; strength 0.93), "response to stimulus" ($n = 5$; strength 0.72), and "regulation of catabolic process" ($n = 4$; strength 1.72), as well as for the cellular components "protein-containing complex" ($n = 4$; strength 0.859) and "extracellular space" ($n = 3$; strength 1.26) (Figure 4).

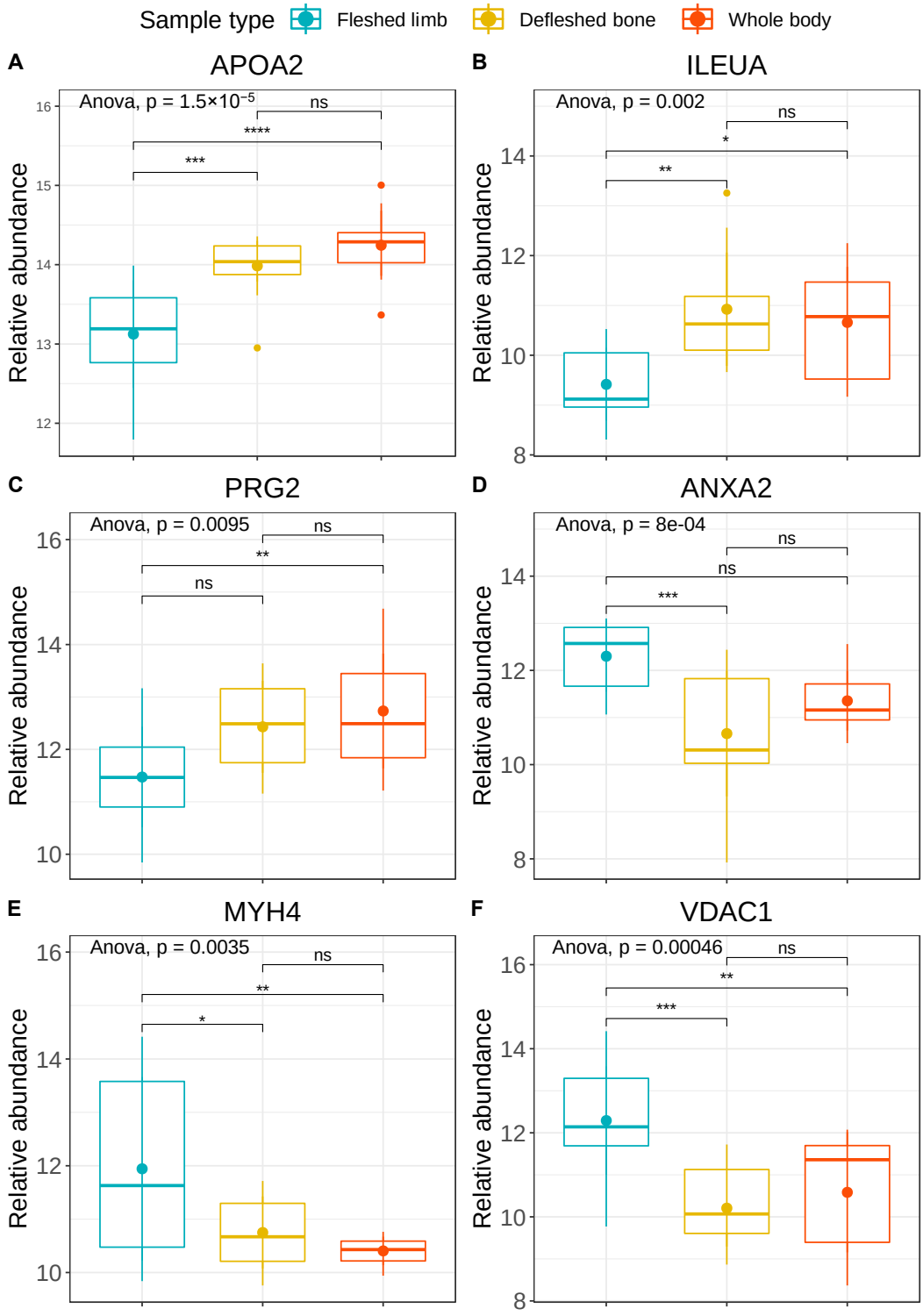


Figure 3. Boxplots with the relative abundances (arcsinh normalised) of (**A**) APOA2, (**B**) ILEUA, (**C**) PRG2, (**D**) ANXA2, (**E**) MYH4, and (**F**) VDAC1. Fleshed exposed limbs are represented in blue, defleshed buried bones in yellow, and exposed whole bodies in red. The plots provide extended *p* values for multivariate ANOVA and coded significance for Tukey's HSD test ($p < 0.0001$ '****'; $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'; $p > 0.05$ 'ns').

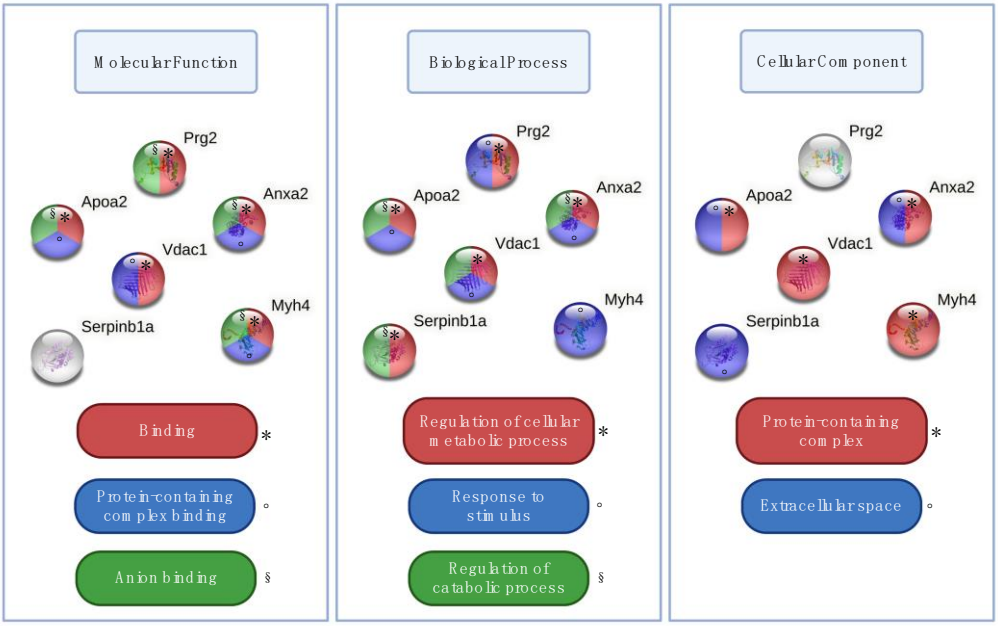


Figure 4. Functional enrichments in GO terms (molecular functions, biological processes, and cellular components) for the six proteins identified as being significant among the groups. For each GO class, different colours and symbols represent a specific GO term.

When examining the pairwise relationships between the different groups, it was clear that fleshed limbs differ from the other two groups. In particular, the abundance of APOA2 (Figure 3A) seemed to be significantly lower in fleshed limbs compared to both defleshed bones and whole bodies. These latter two groups did not show any significant differences between one another. ILEUA2 and PRG2 were also less abundant in fleshed limbs than in the other two sample types, and showed significant differences between fleshed limbs and both defleshed bones and whole bodies (Figure 3B), and between fleshed limbs and whole bodies (Figure 3C). In contrast, ANXA2 and VDAC1 were more abundant in limbs covered in flesh compared to the remaining groups, and the differences were moderately-to-highly significant (Figure 3D,F). Finally, MYH4 showed a significant difference between fleshed limb and whole-body deposition, and a less strong (but still significant) difference between fleshed limbs and defleshed bones (Figure 3E). The same trend also applies when the different type of depositional environment is taken into consideration, as defleshed bones were buried, whereas fleshed limbs and whole bodies were placed on the plastic surface and exposed to environmental factors (such as humidity, insects, environmental microorganisms, etc.).

4. Discussion

This work was aimed at understanding whether significant differences could be found among the proteomes extracted from limb bones deposited/buried into different depositional environments and subjected to different post-mortem conditions, which should have involved different causative agents and impacted the progression of bioerosion and early diagenetic phenomena, over a relatively short period of time ranging from 0 to 28 weeks. In order to do so, we examined the proteomes collected from either exposed whole remains, exposed fleshed back limbs, or buried defleshed bones.

To evaluate which parameters might have contributed the most to the observed proteomic results and associated diagenetic phenomena, it is important to summarize which variables were expected to take place in decomposition in the various sample types and depositional environments tested. Whole bodies exposed to the surface were subjected to the effects of autolysis and to the combined action of both intrinsic bacteria and environmental aerobic bacteria and insects (e.g., blowflies). In this case, the environmental

conditions (e.g., temperature, humidity, rainfall, etc.) have also directly impacted the decomposition rate of the carcasses. A very similar situation can be expected for the fleshed limbs exposed on the surface, with the main difference being that limbs disarticulated from the rest of the body may or may not have been colonised by gut bacteria, depending on the speed at which those bacteria travelled along the vascular system reaching the limbs (and consequently the bones) post-mortem and prior to their dissection. Hence, overall, the effects that internal bacteria had on bone diagenesis and bioerosion should be limited in comparison with those observed on whole carcasses. Finally, the last scenario is the one that differs the most from the other two; in this last case, buried defleshed bones were subjected predominantly to the action of exogenous (e.g., soil) bacteria, soil insects, and autolytic phenomena, and only in a minor way by the presence of intrinsic gut bacteria that eventually managed to reach the bones prior to their disarticulation. Moreover, the environmental conditions affected the decomposition of these samples in a less direct way, due to the surrounding soil matrix that protected them from, for example, drastic temperature changes, direct rainfall, and changes in humidity.

Within this work we did not find significant differences in relation to specific PMIs, but we observed a clear difference between the proteomes of the exposed fleshed limbs and those of the exposed whole remains or buried defleshed bones. In particular, we highlighted six proteins whose abundance was significantly affected by the sample type and by the environmental deposition.

Between the proteomes from the exposed fleshed limbs (subjected predominantly to autolysis and to environmental microorganisms and factors) and from the buried defleshed bones (subjected predominantly to the action of autolysis and soil bacteria), we found APOA2 to be the greatest contributor among the six proteins with different abundances within the various conditions tested. APOA2 is a plasma protein, and was seen to be significantly less abundant in fleshed exposed limbs compared to the other two groups (Figure 3A). No significant differences were shown between buried bones and whole bodies. This suggests that the presence of a cut on the limb (e.g., following the excision of the limb from the body) and the following exposure of the tissues to the external environment allows the bodily fluids (including blood) to flow out faster compared with the whole carcasses. Additionally, bones exposed to external factors (including insects and bacteria, but also environmental conditions) can be subjected to considerably more protein decay than bones either "protected" by an intact carcass or surrounded by soil. It is frequent to find plasma and muscle proteins in bones as residuals from surrounding soft tissues, even when muscles are carefully removed (e.g., in the example of defleshed bones, Procopio and Buckley, data not published). Overall, this could explain the decrease in the number of plasma proteins found in specimens taken from the fleshed limbs group.

These possible interpretations are supported by the fact that the carcasses did not reach the complete skeletonization stage (according to their total body scores in accordance with the table proposed by Adlam and Simmons [63], data not shown) at the end of the experiment, with bone exposure being only <50% of the scored area.

We also noticed similar behaviour for two bone marrow proteins—namely, ILEUA and PRG2—although their contribution to the observed variance was smaller than that found for APOA2. These bone marrow proteins showed similar abundances in the buried bones and in the whole-body samples, but were notably less abundant in the exposed fleshed limbs. This further supports our interpretation that the burial environment and the entire body mass could function as protection from external environmental conditions, reducing the time of exposure of the bone to the mechanical action of rain. These results indicate that bone marrow and plasma proteins are degraded in a faster way in the exposed fleshed limbs in comparison with exposed whole bodies and buried bones, although the reasons for this finding are still not entirely clear, and warrant further investigation. It does not seem illogical that this behaviour is linked to groups of proteins with a lower affinity for hydroxyapatite and bone collagen [17]. In the theoretical absence of soil or gut bacteria, proteins of these groups are mostly degraded by autolysis and bioerosion, while mineral-

binding proteins have shown more longevity and, therefore, represent an ideal target for archaeological research [4,24,46].

MYH4, ANXA2, and VDAC1 were found to be more abundant in the fleshed samples than in the whole ones or in the buried bones, indicating that these proteins may have been attacked by bacteria (either gut or soil bacteria) more than by the extrinsic factors. ANXA2 is a calcium-binding protein involved in osteoclast formation and bone resorption [64]. The higher abundance found in the exposed fleshed limbs could reflect the fact that proteins more intimately connected to hydroxyapatite are less prone to degradation from autolytic processes and environmental factors, but are attacked by the microbial action during bone diagenesis, even in relatively short forensic time frames. Again, for this protein significant differences were only found between fleshed limbs and the remaining two groups, but not between defleshed bones and whole bodies. These findings support what was previously stated for plasma and marrow proteins. The presence of both gut and soil bacteria here plays a major role in reducing the abundance of proteins with high calcium ion affinity. Despite no significance difference being found between the two groups, and so between the two distinct bacterial sources, Figure 3D shows the detrimental effect of microbial action on this protein.

MYH4 and VDAC1 behaved in a similar way to ANXA2. Both of these proteins are highly expressed in the skeletal muscle: myosins are well known for their ATPase activity in the skeletal muscle that allows muscle contraction [65], while VDAC1 is involved in the transport of ATP in the sarcoplasmic reticulum of the skeletal muscles (in addition to mitochondria) [66]. The abundance of these proteins was constantly reduced in the presence of either gut or soil bacteria, as shown by the very small deviations from the median recorded for the defleshed bones and for the whole-body samples. The identification of muscle proteins through bone proteomic analyses is not new or surprising, and has been previously shown where the PMI was not long enough to allow for the complete decomposition of the soft tissues and of the proteins associated with them [13,14]. The fact that muscle proteins were degraded more effectively in the samples where gut bacteria were present (e.g., whole bodies) than in the exposed limbs was expected, since the autolytic processes on muscle tissue normally take place in the first 24–28 h post-mortem [67], and after this period, gut bacteria are the main drivers of additional tissue degradation. Additionally, the low abundance of muscle proteins in buried bones can be explained by the action of soil bacteria combined with the reduced amount of skeletal muscle tissue available at the starting point of the decomposition (after defleshing).

Despite the promising results, the present study is not without limitations. One of these is the use of rats instead of the more commonly used pigs as analogues for humans in decomposition studies. However, it is not uncommon to find forensic studies where small animals (such as rats or mice) have been used instead of pigs in order to increase the sample size and the reproducibility of the results among biological replicates (e.g., same breed, age, food intake, etc.) [38]. Additionally, to further improve the understanding of bone taphonomic phenomena and bioerosion via proteomics, future studies should include more frequent timepoints for sample collection, and could be expanded to prolonged timescales, with an increased number of samples and with the inclusion of an additional indoor/laboratory depositional environment that will exclude the presence of environmental bacteria in order to simplify the model and the interpretation of the results. This will eventually allow for a better interpretation of the interactions between different factors and bone proteomes by implementing more explanatory statistical analyses.

5. Conclusions

Overall, these findings suggest that, despite the fact that bone proteomics does not allow for the distinction between protein degradation caused by different sources of bacteria—such as gut or soil bacteria—it does allow for the discrimination of samples subjected either to bioerosion or to the action of extrinsic factors. This study allowed us to understand how and which NCPs are more degraded in different scenarios, ultimately

providing insights into the survival of bone biomolecules within different conditions. More specifically, the results reported in the present study show that muscle proteins and calcium- and collagen-binding proteins are more prone to degradation by bacterial attack than by hydrolytic and extrinsic processes, even after relatively short timeframes, such as the ones investigated in our study. On the other hand, plasma and bone marrow proteins seem to be protected by the presence of an intact body mass or by the burial environment. To conclude, proteomic analyses show the potential to reveal information that cannot be obtained with more classical approaches regarding taphonomic events occurring in relatively short timeframes, and this should be considered for future studies aimed at better understanding the extent of diagenesis in various conditions, and of the decay of the biomolecules in bones.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biology10060460/s1: Data S1: Protein exports for normalized and raw abundances of the samples grouped by: "Fleshed vs. defleshed vs. whole"; "Fleshed vs. defleshed"; "Fleshed vs. whole"; "Defleshed vs. whole"; "Buried vs. exposed"; and "PMI". Figure S1: Monthly temperatures collected at the HuddersFIELD site and average monthly rainfall at the HuddersFIELD site during the field experiment.

Author Contributions: Conceptualization, C.A.M. and A.W.; data curation, N.P. and A.B.; formal analysis, N.P.; funding acquisition, N.P.; investigation, N.P., C.A.M., S.S., and A.W.; methodology, N.P.; project administration, N.P. and A.W.; resources, A.B. and A.W.; supervision, A.W.; visualization, N.P. and A.B.; writing—original draft, N.P., C.A.M., and A.B.; writing—review and editing, N.P., C.A.M., A.B., and A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Royal Society [Research Grant (N.P.) under grant RGS/R1/191371], and by the UKRI [Future Leaders Fellowship (N.P. and A.B.) under grant MR/S032878/1].

Institutional Review Board Statement: The study was conducted in accordance with the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Huddersfield (School Research Integrity and Ethics Committee—grant number—SAS REIC 18-2611-1).

Informed Consent Statement: Not applicable.

Data Availability Statement: Normalised and raw abundances for the proteins analysed in this study can be all found in the Supplementary Data S1. The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [68] partner repository with the dataset identifier PXD026042.

Acknowledgments: The graphical abstract and Figure 4 have been created with BioRender.com.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. White, L.; Booth, T.J. The origin of bacteria responsible for bioerosion to the internal bone microstructure: Results from experimentally-deposited pig carcasses. *Forensic Sci. Int.* **2014**, *239*, 92–102. [CrossRef]
2. Collins, C.J.; Kozyrev, M.; Frank, M.; Andriotis, O.G.; Byrne, R.A.; Kiener, H.P.; Pretterklieber, M.L.; Thurner, P.J. The impact of age, mineralization, and collagen orientation on the mechanics of individual osteons from human femurs. *Materialia* **2020**, *9*, 100573. [CrossRef]
3. Jans, M.M.E.; Kars, H.; Nielsen-Marsh, C.M.; Smith, C.I.; Nord, A.G.; Arthur, P.; Earl, N. In situ preservation of archaeological bone: A histological study within a multidisciplinary approach. *Archaeometry* **2002**, *44*, 343–352. [CrossRef]
4. Trueman, C.N.; Martill, D.M. The long-term survival of bone: The role of bioerosion. *Archaeometry* **2002**, *44*, 371–382. [CrossRef]
5. Turner-Walker, G. Light at the end of the tunnels? The origins of microbial bioerosion in mineralised collagen. *Palaeogeogr. Palaeoclim. Palaeoecol.* **2019**, *529*, 24–38. [CrossRef]
6. Jans, M.; Nielsen-Marsh, C.; Smith, C.; Collins, M.; Kars, H. Characterisation of microbial attack on archaeological bone. *J. Archaeol. Sci.* **2004**, *31*, 87–95. [CrossRef]

7. Brown, S.; Higham, T.; Slon, V.; Pääbo, S.; Meyer, M.; Douka, K.; Brock, F.; Comeskey, D.; Procopio, N.; Shunkov, M.; et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **2016**, *6*, 23559. [CrossRef] [PubMed]
8. Meyer, M.; Kircher, M.; Gansauge, M.-T.; Li, H.; Racimo, F.; Mallick, S.; Schraiber, J.G.; Jay, F.; Prüfer, K.; De Filippo, C.; et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **2012**, *338*, 222–226. [CrossRef]
9. Buckley, M.; Anderung, C.; Penkman, K.; Raney, B.J.; Götherström, A.; Thomas-Oates, J.; Collins, M.J. Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites. *J. Archaeol. Sci.* **2008**, *35*, 1756–1764. [CrossRef]
10. Buckley, M. Ancient collagen reveals evolutionary history of the endemic South American 'ungulates'. *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 20142671. [CrossRef]
11. Prieto-Bonete, G.; Pérez-Cárceles, M.D.; Maurandi-López, A.; Pérez-Martínez, C.; Luna, A. Association between protein profile and postmortem interval in human bone remains. *J. Proteom.* **2019**, *192*, 54–63. [CrossRef] [PubMed]
12. Pérez-Martínez, C.; Pérez-Cárceles, M.D.; Legaz, I.; Prieto-Bonete, G.; Luna, A. Quantification of nitrogenous bases, DNA and Collagen type I for the estimation of the postmortem interval in bone remains. *Forensic Sci. Int.* **2017**, *281*, 106–112. [CrossRef] [PubMed]
13. Procopio, N.; Williams, A.; Chamberlain, A.T.; Buckley, M. Forensic proteomics for the evaluation of the post-mortem decay in bones. *J. Proteom.* **2018**, *177*, 21–30. [CrossRef] [PubMed]
14. Mizukami, H.; Hathway, B.; Procopio, N. Aquatic Decomposition of Mammalian Corpses: A Forensic Proteomic Approach. *J. Proteome Res.* **2020**, *19*, 2122–2135. [CrossRef] [PubMed]
15. Nielsen-Marsh, C. Biomolecules in fossil remains: Multidisciplinary approach to endurance. *Biochemist* **2002**, *24*, 12–14. [CrossRef]
16. Nielsen-Marsh, C.M.; Hedges, R.E.; Mann, T.; Collins, M.J. A preliminary investigation of the application of differential scanning calorimetry to the study of collagen degradation in archaeological bone. *Thermochim. Acta* **2000**, *365*, 129–139. [CrossRef]
17. Buckley, M.; Wadsworth, C. Proteome degradation in ancient bone: Diagenesis and phylogenetic potential. *Palaeogeogr. Palaeoclim. Palaeoecol.* **2014**, *416*, 69–79. [CrossRef]
18. Sawafuji, R.; Cappellini, E.; Nagaoka, T.; Fotakis, A.K.; Jersie-Christensen, R.R.; Olsen, J.V.; Hirata, K.; Ueda, S. Proteomic profiling of archaeological human bone. *R. Soc. Open Sci.* **2017**, *4*, 161004. [CrossRef]
19. Wadsworth, C.; Procopio, N.; Anderung, C.; Carretero, J.-M.; Iriarte, E.; Valdiosera, C.; Elburg, R.; Penkman, K.; Buckley, M. Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. *J. Proteom.* **2017**, *158*, 1–8. [CrossRef]
20. Collins, M.J.; Riley, M.S.; Child, A.M.; Turner-Walker, G. A Basic Mathematical Simulation of the Chemical Degradation of Ancient Collagen. *J. Archaeol. Sci.* **1995**, *22*, 175–183. [CrossRef]
21. Hedges, R.E.M. Bone diagenesis: An overview of processes. *Archaeometry* **2002**, *44*, 319–328. [CrossRef]
22. Jans, M.M.E. Microbial bioerosion of bone—A review. In *Current Developments in Bioerosion*; Wisshak, M., Tapanila, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 397–413. [CrossRef]
23. Collins, M.; Nielsen-Marsh, C.M.; Hiller, J.; I Smith, C.; Roberts, J.P.; Prigodich, R.V.; Wess, T.J.; Csapo, J.; Millard, A.R.; Turner-Walker, G. The survival of organic matter in bone: A review. *Archaeometry* **2002**, *44*, 383–394. [CrossRef]
24. Wadsworth, C.; Buckley, M. Proteome degradation in fossils: Investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 605–615. [CrossRef] [PubMed]
25. Masters, P.M. Preferential preservation of noncollagenous protein during bone diagenesis: Implications for chronometric and stable isotopic measurements. *Geochim. Cosmochim. Acta* **1987**, *51*, 3209–3214. [CrossRef]
26. Smith, C.; Craig, O.; Prigodich, R.; Nielsen-Marsh, C.; Jans, M.; Vermeer, C.; Collins, M. Diagenesis and survival of osteocalcin in archaeological bone. *J. Archaeol. Sci.* **2005**, *32*, 105–113. [CrossRef]
27. Creamer, J.I.; Buck, A.M. The assaying of haemoglobin using luminol chemiluminescence and its application to the dating of human skeletal remains. *Lumin* **2009**, *24*, 311–316. [CrossRef] [PubMed]
28. Costa, I.; Carvalho, F.; Magalhães, T.; De Pinho, P.G.; Silvestre, R.; Dinis-Oliveira, R.J. Promising blood-derived biomarkers for estimation of the postmortem interval. *Toxicol. Res.* **2015**, *4*, 1443–1452. [CrossRef]
29. Takeichi, S.; Tokunaga, I.; Yoshima, K.; Maeiwa, M.; Bando, Y.; Kominami, E.; Katunuma, N. Mechanism of postmortem autolysis of skeletal muscle. *Biochem. Med.* **1984**, *32*, 341–348. [CrossRef]
30. Lametsch, R.; Karlsson, A.; Rosenvold, K.; Andersen, H.J.; Roepstorff, P.; Bendixen, E. Postmortem Proteome Changes of Porcine Muscle Related to Tenderness. *J. Agric. Food Chem.* **2003**, *51*, 6992–6997. [CrossRef]
31. Carter, D.O. The importance of microbial communities in the estimation of the time since death. In *Estimation of the Time since Death: Current Research and Future Trends*; Hayman, J., Oxenham, M., Eds.; Academic Press: London, UK, 2020; pp. 109–139. [CrossRef]
32. Melvin, J.R.; Cronholm, L.S.; Simson, L.R.; Isaacs, A.M. Bacterial transmigration as an indicator of time of death. *J. Forensic Sci.* **1984**, *29*, 412–417. [CrossRef]
33. Kellerma, G.D.; Waterman, N.G.; Scharfenberger, L.F. Demonstration in vitro of postmortem bacterial transmigration. *Am. J. Clin. Pathol.* **1976**, *66*, 911–915. [CrossRef] [PubMed]
34. Hyde, E.R.; Haarmann, D.P.; Lynne, A.M.; Bucheli, S.R.; Petrosino, J.F. The living dead: Bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. *PLoS ONE* **2013**, *8*, e77733. [CrossRef]

35. Finley, S.J.; Benbow, M.E.; Javan, G.T. Microbial communities associated with human decomposition and their potential use as postmortem clocks. *Int. J. Leg. Med.* **2014**, *129*, 623–632. [CrossRef] [PubMed]
36. Damann, F.E.; Jans, M.M.E. Microbes, anthropology, and bones. In *Forensic Microbiology*; Carter, D.O., Tomberlin, J.K., Benbow, M.E., Metcalf, J.L., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2017; p. 319.
37. Hyde, E.R.; Haarmann, D.P.; Petrosino, J.F.; Lynne, A.M.; Bucheli, S.R. Initial insights into bacterial succession during human decomposition. *Int. J. Leg. Med.* **2014**, *129*, 661–671. [CrossRef]
38. Metcalf, J.L.; Parfrey, L.W.; Gonzalez, A.; Lauber, C.L.; Knights, D.; Ackermann, G.; Humphrey, G.C.; Gebert, M.J.; Van Treuren, W.; Berg-Lyons, D.; et al. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *eLife* **2013**, *2*, e01104. [CrossRef]
39. Cobaugh, K.L.; Schaeffer, S.M.; DeBruyn, J.M. Functional and structural succession of soil microbial communities below decomposing human cadavers. *PLoS ONE* **2015**, *10*, e0130201. [CrossRef]
40. Damann, F.E.; Williams, D.E.; Layton, A.C. Potential use of bacterial community succession in decaying human bone for estimating postmortem interval. *J. Forensic Sci.* **2015**, *60*, 844–850. [CrossRef]
41. Kontopoulos, I.; Nystrom, P.; White, L. Experimental taphonomy: Post-mortem microstructural modifications in *Sus scrofa* domesticus bone. *Forensic Sci. Int.* **2016**, *266*, 320–328. [CrossRef]
42. Brönnimann, D.; Portmann, C.; Pichler, S.L.; Booth, T.J.; Röder, B.; Vach, W.; Schibler, J.; Rentzel, P. Contextualising the dead—Combining geoarchaeology and osteo-anthropology in a new multi-focus approach in bone histotaphonomy. *J. Archaeol. Sci.* **2018**, *98*, 45–58. [CrossRef]
43. Reiche, I.; Favre-Quattropani, L.; Vignaud, C.; Bocherens, H.; Charlet, L.; Menu, M. A multi-analytical study of bone diagenesis: The Neolithic site of Bercy (Paris, France). *Meas. Sci. Technol.* **2003**, *14*, 1608–1619. [CrossRef]
44. Fierer, N.; Schimel, J.P.; A Holden, P. Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* **2003**, *35*, 167–176. [CrossRef]
45. Nielsen-Marsh, C.M.; Hedges, R.E. Patterns of Diagenesis in Bone I: The Effects of Site Environments. *J. Archaeol. Sci.* **2000**, *27*, 1139–1150. [CrossRef]
46. Goff, M.L. Early post-mortem changes and stages of decomposition in exposed cadavers. *Exp. Appl. Acarol.* **2009**, *49*, 21–36. [CrossRef]
47. Langley, N.R.; Tersigni-Tarrant, M.A. (Eds.) *Forensic Anthropology. A Comprehensive Introduction*; CRC Press: Boca Raton, FL, USA, 2017.
48. Fernández-Jalvo, Y.; Andrews, P.; Pesquero, D.; Smith, C.; Marín-Monfort, D.; Sánchez, B.; Geigl, E.-M.; Alonso, A. Early bone diagenesis in temperate environments. *Palaeogeogr. Palaeoclim. Palaeoecol.* **2010**, *288*, 62–81. [CrossRef]
49. Yoshino, M.; Kimijima, T.; Miyasaka, S.; Sato, H.; Seta, S. Microscopical study on estimation of time since death in skeletal remains. *Forensic Sci. Int.* **1991**, *49*, 143–158. [CrossRef]
50. Hyde, E.R.; Metcalf, J.L.; Bucheli, S.R.; Lynne, A.M.; Knight, R. Microbial communities associated with decomposing corpses. In *Forensic Microbiology*; Wiley Online Books; John Wiley Sons Ltd.: Chichester, UK, 2017; pp. 245–273.
51. Li, D.; Zhu, Z.; Sun, D.-W. Effects of freezing on cell structure of fresh cellular food materials: A review. *Trends Food Sci. Technol.* **2018**, *75*, 46–55. [CrossRef]
52. Speck, M.L.; Ray, B. Effects of Freezing and Storage on Microorganisms in Frozen Foods: A Review. *J. Food Prot.* **1977**, *40*, 333–336. [CrossRef] [PubMed]
53. Procopio, N.; Ghignone, S.; Williams, A.; Chamberlain, A.; Mello, A.; Buckley, M. Metabarcoding to investigate changes in soil microbial communities within forensic burial contexts. *Forensic Sci. Int. Genet.* **2019**, *39*, 73–85. [CrossRef]
54. Procopio, N.; Buckley, M. Minimizing Laboratory-Induced Decay in Bone Proteomics. *J. Proteome Res.* **2017**, *16*, 447–458. [CrossRef] [PubMed]
55. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]
56. Huber, W.; Von Heydebreck, A.; Sültmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **2002**, *18*, S96–S104. [CrossRef]
57. Blair, H.C.; Kalyvioti, E.; Papachristou, I.N.; Tourkova, I.L.; A Syggelos, S.; Deligianni, D.; Orkoula, M.G.; Kontoyannis, C.G.; A Karavia, E.; Kypreos, E.K.; et al. Apolipoprotein A-1 regulates osteoblast and lipoblast precursor cells in mice. *Lab. Investig.* **2016**, *96*, 763–772. [CrossRef] [PubMed]
58. Fouret, P.; Du Bois, R.M.; Bernaudin, J.F.; Takahashi, H.; Ferrans, V.J.; Crystal, R.G. Expression of the neutrophil elastase gene during human bone marrow cell differentiation. *J. Exp. Med.* **1989**, *169*, 833–845. [CrossRef]
59. Overgaard, M.T.; Oxvig, C.; Christiansen, M.; Lawrence, J.B.; Conover, C.A.; Gleich, G.J.; Sottrup-Jensen, L.; Haaning, J. Messenger Ribonucleic Acid Levels of Pregnancy-Associated Plasma Protein-A and the Proform of Eosinophil Major Basic Protein: Expression in Human Reproductive and Nonreproductive Tissues1. *Biol. Reprod.* **1999**, *61*, 1083–1089. [CrossRef] [PubMed]
60. Hopwood, B.; Tsykin, A.; Findlay, D.; Fazzalari, N. Gene expression profile of the bone microenvironment in human fragility fracture bone. *Bone* **2009**, *44*, 87–101. [CrossRef] [PubMed]
61. Anflous-Pharayra, K.; Cai, Z.-J.; Craigen, W.J. VDAC1 serves as a mitochondrial binding site for hexokinase in oxidative muscles. *Biochim. Biophys. Acta BBA Bioenerg.* **2007**, *1767*, 136–142. [CrossRef]

62. Yuan, H.; Niu, Y.; Liu, X.; Yang, F.; Niu, W.; Fu, L. Proteomic Analysis of Skeletal Muscle in Insulin-Resistant Mice: Response to 6-Week Aerobic Exercise. *PLoS ONE* **2013**, *8*, e53887. [CrossRef] [PubMed]
63. Adlam, R.E.; Simmons, T. The Effect of Repeated Physical Disturbance on Soft Tissue Decomposition?Are Taphonomic Studies an Accurate Reflection of Decomposition? *J. Forensic Sci.* **2007**, *52*, 1007–1014. [CrossRef]
64. Genetos, D.C.; Wong, A.; Watari, S.; Yellowley, C.E. Hypoxia increases Annexin A2 expression in osteoblastic cells via VEGF and ERK. *Bone* **2010**, *47*, 1013–1019. [CrossRef]
65. Bárány, M. ATPase Activity of Myosin Correlated with Speed of Muscle Shortening. *J. Gen. Physiol.* **1967**, *50*, 197–218. [CrossRef]
66. Shoshan-Barmatz, V.; Hadad, N.; Feng, W.; Shafir, I.; Orr, I.; Varsanyi, M.; Heilmeyer, L.M. VDAC/porin is present in sarcoplasmic reticulum from skeletal muscle. *FEBS Lett.* **1996**, *386*, 205–210. [CrossRef]
67. Pittner, S.; Ehrenfellner, B.; Monticelli, F.C.; Zissler, A.; Sänger, A.M.; Stoiber, W.; Steinbacher, P. Postmortem muscle protein degradation in humans as a tool for PMI delimitation. *Int. J. Leg. Med.* **2016**, *130*, 1547–1555. [CrossRef] [PubMed]
68. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef] [PubMed]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Biology Editorial Office
E-mail: biology@mdpi.com
www.mdpi.com/journal/biology

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com

ISBN 978-3-0365-5431-0