UNIVERSIDADE D
COIMBRA

David Senhora Navega

MULTIFACTORIAL SKELETAL AGE ESTIMATION IN FORENSIC ANTHROPOLOGY AND MEDICINE
A MACHINE LEARNING APPROACH

Tese no âmbito do Doutoramento em Antropologia, ramo de especialização em Antropologia Forense, orientada pela Professora Doutora Eugénia Maria Guedes Pinto Antunes da Cunha e pelo Professor Doutor Ernesto Jorge Fernando Costa e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Dezembro de 2022

UNIVERSIDADE D
COIMBRA

David Senhora Navega

MULTIFACTORIAL SKELETAL AGE ESTIMATION IN FORENSIC ANTHROPOLOGY AND MEDICINE

A MACHINE LEARNING APPROACH

Doctoral thesis submitted in partial fulfillment of the Doctoral Program in Anthropology, specialization in Forensic Anthropology, supervised by Full Professor Eugénia Maria Guedes Pinto Antunes da Cunha and Full Professor Ernesto Jorge Fernando Costa and presented to the Department of Life Sciences, Faculty of Science and Technology, University of Coimbra

December 2022

Financial support by Fundação para a Ciência e Tecnologia.

SFRH/BD/99676/2014.

Multifactorial skeletal age estimation in forensic anthropology and medicine: a machine learning approach

All models are wrong, but some are useful.

George Box

[*This page has been intentionally left blank*]

Abstract

Age-at-death assessment is a crucial step in the identification process of human skeletal remains. Nonetheless, in adult individuals this task is particularly difficult to achieve accurately due to the variability of the senescence processes. The literature argues in favor of a multifactorial approach to skeletal age assessment to obtain precise estimates. Conceptually a multifactorial perspective can be argued as the most effective approach because skeletal traits display different age-related trajectories and onsets. However, adult age estimation struggles with methodological inconsistencies. Techniques that use multiple skeletal indicators are often limited to the cranial sutures and the pelvic joints. More generic procedures for multifactorial analysis have also been proposed, but with poor adoption in forensic casework because they require seriation or advanced mathematical knowledge to be put into action.

The present thesis aimed to lay a foundation to tackle some of the challenges of macroscopic adult skeletal age estimation, especially in its holistic or multifactorial aspect. The main objective of this work was to propose a new method for multifactorial age estimation using an interdisciplinary approach bridging anthropology and computer science. From an anthropological perspective, a novel macroscopic technique for skeletal analysis was developed. This proposal incorporates a total of 64 skeletal traits covering major joints and musculoskeletal complexes, integrating well established age-related markers with less explored ones. A dataset comprising information on 500 identified skeletons was used to establish a reference dataset (19–101 years old, 250 males and 250 females) for adult age estimation. A computational framework based on machine learning using randomized deep neural networks was implemented and validated. This approach tackled age estimation from a function approximation perspective as regression problem to infer both point and prediction interval estimates.

Two experiments were conducted computationally to assess the value of the multifactorial approach: the first experiment compared multi-trait or multifactorial models against classic models using specific anatomical regions or skeletal traits only; the second experiment assessed the accuracy of age estimation from fractioned multifactorial models using randomly chosen traits.

Based on cross-validation analysis, results demonstrate that age estimation from skeletal remains can be accurately inferred across the entire adult age span, approximately with 6 years mean absolute error. Informative estimates and prediction intervals can be obtained for the elderly population. Multifactorial models introduce a two-to-six-fold reduction in the mean absolute error and prediction bias compared to standard models. Virtually every combination of random traits resulted in models with comparable or better performance than the models built of specific anatomic regions as traditionally encounter in macroscopic age estimation methods. This finding supports the value of multifactorial age estimation over methods the focus solely on a single anatomical structure.

A novel software, DRNNAGE, was built to operationalized and integrate the new method proposed in this thesis, providing an intuitive interface and freely distributed under an open-source license.

Keywords

Forensic anthropology; Skeletal age; Age estimation; Artificial intelligence; Machine learning

Resumo

A estimativa da idade à morte é uma etapa crucial no processo de identificação de restos humanos esqueléticos. No entanto, em indivíduos adultos, a sua avaliação é particularmente difícil de ser realizada com precisão devido à variabilidade dos processos de senescência.

A literatura defende uma abordagem multifatorial para avaliação da idade esquelética de modo a obter estimativas mais exatas e precisas. Esta perspetiva pode ser conceptualmente argumentada como a mais eficaz, porque os diversos marcadores esqueléticos e a sua relação com a idade exibem diferentes trajetórias. Todavia, na estimativa da idade à morte em adultos prevalecem inconsistências metodológicas, uma vez que as técnicas que usam diversos marcadores esqueléticos resumem-se às suturas cranianas e às articulações pélvicas. Têm sido sugeridos procedimentos mais genéricos, mas com pouca utilização em contexto pericial porque requerem técnicas de seriação ou conhecimento matemático avançado.

A presente tese procurou solucionar alguns dos problemas da estimativa da idade em adultos por meio de análise macroscópica tendo em conta diversos marcadores osteológicos. O principal objetivo foi desenvolver um novo método para estimativa de idade à morte em adultos usando uma abordagem interdisciplinar que une a antropologia forense e a inteligência artificial. Do ponto de vista antropológico foi proposta uma nova técnica para avaliação macroscópica que incorpora um total de 64 marcadores esqueléticos que cobrem as principais articulações e complexos músculo-esqueléticos, integrando marcadores bem estabelecidos com outros menos explorados.

Para estabelecer um novo conjunto de dados de referência, foi estudada uma amostra composta por 500 esqueletos identificados provenientes de duas coleções osteológicas da Universidade de Coimbra (19-101 anos, 250 homens e 250 mulheres). Neste

trabalho foi implementada e validada uma nova abordagem computacional tendo por base técnicas de inteligência artificial, concretamente redes neuronais artificiais profundas aleatorizadas. Nesta metodologia a estimativa da idade é tratada como um problema de regressão para a estimativa pontual e intervalar. Foram assim conduzidas computacionalmente duas experiências para testar o seu valor: na primeira, comparam-se modelos multivariados com modelos clássicos usando apenas regiões anatómicas específicas; na segunda avaliou-se a precisão da estimativa de idade a partir de modelos multivariados fracionados, que usam apenas uma parte das características esqueléticas escolhidas aleatoriamente.

Os resultados com base na análise de validação cruzada, demonstram que a estimativa de idade a partir de remanescentes osteológicos pode ser inferida com precisão em toda a faixa etária adulta incluindo indivíduos com idade muito avançada, reduzindo erro médio absoluto de para seis anos aproximadamente.

Praticamente todas as combinações aleatórias de marcadores ósseos resultaram em modelos com desempenho comparável ou superior ao dos modelos construídos de regiões anatómicas específicas, diminuindo em duas a seis vezes o erro médio absoluto e o viés de estimativa em comparação com os modelos padrão. Estes resultados reforçam a importância de uma análise multifatorial na estimativa da idade em adultos.

Foi desenvolvido um novo software, DRNNAGE, para implementar o método supramencionado. O software possui uma interface intuitiva e é distribuído gratuitamente sob uma licença de código aberto.

Palavras-Chave

Antropologia forense; Idade esquelética; Estimativa da idade à morte; Inteligência artificial; Aprendizagem automática

Acknowledgments*
Agradecimentos

Começo por agradecer aos meus orientadores, a Professora Eugénia Cunha e o Professor Ernesto Costa. A ambos agradeço terem aceite orientar este trabalho de natureza interdisciplinar e toda a liberdade que me concederam na definição e materialização do que este trabalho de investigação viria a ser. Agradeço-lhes a ajuda em todas as dúvidas e desafios, científicos e pessoais, que tiveram que ser ultrapassados. Para sempre grato por nunca terem deixado de acreditar em mim e neste trabalho.

Um agradecimento a todos os membros do Laboratório de Antropologia Forense. Em particular, à Teresa, ao Francisco e ao David que ao longo dos anos desempenharam um papel de mentores e amigos, e que as muitas oportunidades que me proporcionaram claramente ajudaram a traçar o meu percurso académico e profissional. Um muito obrigado à equipa do Gabinete Médico-Legal e Forense do Baixo Vouga por todo o apoio e pelas inúmeras vezes que me questionaram quando é que esta etapa chegava ao fim.

Obrigado aos meus pais e ao meu irmão pelo apoio incondicional, o vosso amor e humor foram fundamentais para lidar com os momentos mais difíceis. Um agradecimento muito especial à Daniela Pereira, minha namorada. Sem ti, muito do que está neste documento não seria possível e estaria num perpétuo estado de *quase-quase* acabado. Obrigado pelo teu apoio e carinho, por celebrares as minhas vitórias e me ajudares a relativizar e ultrapassar as minhas frustrações. Obrigado pela tua leitura atenta mesmo perante tantos *gatafunfos*. Obrigado a todos os meus amigos pelos momentos de convívio, em especial ao João Coelho, ao Tó Zé e ao Calil que vezes sem conta ouviram os meus devaneios e lamúrias. Sem a vossa ajuda este o trabalho não teria sido possível.

---

* For personal reasons, acknowledgements are written in Portuguese.

[*This page has been intentionally left blank*]

List of Tables

This page has been intentionally left blank]

List of Figures

[*This page has been intentionally left blank*]

Table of Contents

[*This page has been intentionally left blank*]

1|Introduction: problem statement

Anthropology has emerged as a major force in the forensic sciences. Forensic anthropology is no longer defined solely as a subfield (or application) of biological anthropology. Over the last thirty years, there has been a profound change, a true paradigm shift, and forensic anthropology has evolved into a discipline with its own theoretical and conceptual corpus, as well as a research agenda. Forensic anthropology and its attributions have changed dramatically. Indeed, this evolution has been so marked and dramatic that even some of the most experienced and long-term practicing anthropologists may struggle to conceptualize or be fully proficient in the many areas covered by the discipline nowadays [1, 2].

Although advancement and transformation are fundamental concepts in forensic anthropology, they are not always visible or implemented. For example, in the 1970s, when the field was beginning to gain recognition among forensic sciences and the general public, little to no research was conducted with a strict emphasis on forensic anthropology [1–3]. At the time, anthropologists would conduct skeletal analyses using the classical methods of biological anthropology developed at the turn of the century.

In terms of research and development, the field experienced a true period of inertia. Around the 1980s, the discipline underwent an important period of introspection, recognizing its strengths, weaknesses, and future challenges. In an almost prophetic manner İşcan [3] wrote that forensic anthropology could *"(…) stagnate or even self-destruct if direction of future research is not carefully planned (…)*". Fortunately, forensic anthropology did not implode, instead a tremendous shift in its paradigm took place. Understanding what caused such change is fundamental to contextualize any research effort in the discipline. Change within the discipline flourished from a variety of external and

internal factors. Dirkmaat and collaborators [1, 2] identify the rapid growth and development of DNA analysis, and the several federal court rulings and USA institutional-mandated assessments of the state-of-art of forensic sciences as the most significant external forces changing and shaping the course of forensic anthropology.

It is critical to comprehend the significance and impact of DNA analysis in forensic anthropology. The primary importance of forensic genetics stems from a common goal: identification. The precision of molecular analysis transformed the human identification process, and it is now the undisputed gold standard for this purpose. It is critical to remember that advancements in forensic genetics were not simplistic refinements of its key methodologies. Instead, massive advances in this field have occurred in recent decades, with each stage of its evolution removing critical methodological and technological barriers. These reasons highlight the significance of DNA analysis in shaping the evolution of forensic anthropology as it broadened its goals, attributions, and theoretical framework. Forensic genetics has undoubtedly exerted selective and evolutionary pressure on forensic anthropology. The discipline was forced to choose between improving its own methods and techniques or extinction, because its primary goal, identification, could be achieved with far greater, if not incomparable, accuracy by forensic genetics. Despite this somewhat pessimistic outlook, collaboration with forensic geneticists increased casework for anthropology experts, and this collaborative approach improved both disciplines' results and impact.

Several legal cases trialed in the United States of America courts of law quickly exposed severe limitations in forensic sciences operating procedures. Independent and external analysis of the state-of-art of forensic sciences, the most recent conducted by the prestigious National Academy of Sciences [4], noted several flaws in terms of standardization of procedures and training of the professionals in the forensic sciences community. American supreme court-rulings such as Daubert v. Merrill Dow

Pharmaceuticals, Inc. (1993) and Kumho Tire, Ltd v. Carmichael (1999), certainly enforced best practices in forensic sciences and shaped research agendas, forensic anthropology was not left out. More details on the these court-rulings can be found in Christensen & Crowder [5] and Grivas and Komar [6].

These external factors prompted forensic anthropology to conduct a critical self-evaluation, resulting in a constant redefinition of its objectives and attributions. The recognition of the discipline's inherent limitations was a key development, particularly in an area most classical and defining of the field itself: biological profiling, or the estimation of sex, age, ancestry, and stature of human remains (especially in cases of skeletonization).

The technical shortcomings of biological profiling techniques were quickly recognized by anthropologists. The inadequacy of such methods and techniques was primarily a two-headed problem: the samples used to develop the procedures did not accurately represent the ever-changing biology and demography of individuals presented in forensic cases, and the mathematical procedures underlying them did not fully utilize or were inappropriate for such data in the worst-case scenario. Age-at-death was a component of the biological profile where such issues were more visible and pervasive.

1.1|Scope

Age estimation is an overarching topic with relevance to diverse disciplines such as anthropology, archaeology, demography, genetics, dentistry, pediatrics, gerontology and forensics often convoluted and characterized by significant interdisciplinary dependencies. When addressing this topic is often made a distinction between chronological and biological age [7]. Chronological age is the time passed or years lived from the moment birth, and it is what legally documented for an individual. Biological age is not fully characterized by the time passed but accounts for the physiological status of the individual. While it is

believed that biological age better reflects the true age of an organism or individual, it is important to note that it is a loosely used conceptual artifact that lacks a precise definition [8–12]. Biological age is established through estimation, where age-related biomarkers are used to infer chronological age via the application of regression algorithm [9], the estimate obtained with this process is considered the biological age and it has important applications in public health and preventive medicine [8, 9]. From this formulation it can be inferred that biological age can be unfold in many dimensions depending on the biomarkers used, metabolic biomarkers reflecting metabolic age, dental markers reflecting dental age and so on. The precision and accuracy of biological age will by a product of the age-related biomarkers and the mathematical algorithm used in its estimation.

Within the attributions of anthropology and its application to the forensic context, there is an important division regarding to the state of the subject(s) to whom age estimation is concerned: age estimation pertains and is required both the living and the deceased [13]. Age estimation of the living is out of the scope of this thesis, a detailed overview can be found in Black *et al.* [14]. However, it is important to note that its goal is to solve legal issues relating to minors in respect to imputability or their depiction pedopornographic material, and for adults legal issues regarding to pensionable age and other questions for those lacking valid identification documents. For the dead, age estimation integrates the identification process through the creation of the biological profile of the deceased. In both situations methodological choices to better approach age estimation are, in first place, dependent of the maturational status of the subject – is it a subadult (birth to 20 years old) or an adult (from 20 years old onward).

Age estimation of subadults is performed by assessing skeletal and dental development parameters such as bone growth and maturation, and tooth calcification and eruption. It is widely acknowledged that the growth and development process is complex, with variation introduced at the level of the individual and in response to extrinsic factors

such as nutrition, biological and psychological stress, disease, and socioeconomic status [15]. However, such complexity is much more amenable compared to the aging and senescence process underlying the adult age estimation and is a main reason why subadult age estimation is regularly stated to be more accurate. As aging as a research topic evolve in depth and breadth, its definition has varied [16]. Within the scope of this thesis, (skeletal) aging can be understood as a set of cumulative and progressive functional and structural changes that are manifest after skeletal maturation. Genetics, endocrine function, disease, joint biomechanics, diet, body size, body composition, activity level, and even climate may explain differences in the timing and progression of age-related traits both at inter and intra-individual level [17–19], rendering age estimation from skeletal remains in adults a difficult task.

Adult age estimation evolved immensely over the last 150 years as reflected in diverse array methods and modalities that are available nowadays [20, 21]. Age-at-death can be inferred using biochemical and epigenetic [22–29], histological [30], radiological and densitometric [31–35] and dental [36–40] approaches. Nonetheless, macroscopic techniques are the most used in current forensic settings when addressing identification from human skeletal remains [41, 42]. Macroscopic analysis contrasts with most of the other approaches due to its inexpensive nature as it does not require specialized equipment, reagents, or facilities. However, it does present several limitations and challenges which addressing represents the scope of this thesis.

1.1|Motivation

The main motivation for this thesis is the critical role of age-at-death assessment in the task of biological profile estimation from human remains, as well as its challenging nature and the need for methodological improvement.

In forensic contexts, where achieving identification is pivotal, age-at-death estimation is critical. In the identification of human remains, age-at-death is a major screening factor that assist in narrowing the universe of possible matches; thus, an estimate of this biological parameter is a common request from law enforcement and judicial entities [43]. This process is based on a meticulous examination of skeletal and dental structures with association to age-at-death. The efficiency of age estimation methods exhibits a directional bias, with error increasing with age. Estimating this parameter in neonatal, juvenile, and adolescent remains is highly accurate. Hormonal and genetic mechanisms regulate growth. As a result, the age of subadults follows a more consistent pattern of association with skeletal and dental traits. As growth slows and eventually ceases, age estimation relies on the skeletal tissue response to degeneration through bone and dental remodeling and is thus less accurate due to the wide variation in such processes.

Skeletal age estimation in adult remains is particularly difficult because skeletal morphology and chronological age frequently show a weak relationship (e.g., cranial suture obliteration), and confounding factors such as sex, ancestry, or intra-personal variation all have an impact on skeletal morphology and its age-related expression. Most commonly used adult skeletal age estimation methods rely on macroscopic analysis [41], which means that age-related information is extracted by a human expert through visual and manual examination of the skeletal morphology. This leaves room for subjectivity, adding another thick layer of complexity to an already difficult task.

Despite being a topic where much research was performed in the last years, skeletal age estimation of adult remains still presents many unanswered questions and challenges, especially for the elderly. How to handle age-at-death estimation using multiple morphology indicators are among the problems most commonly identified for which a satisfactory solution has not yet been presented and research is required [41, 43–49]. Also, computational and statistical methods employed in the creation of age estimation

techniques has been a topic of debate and contention [32, 50–62]. The challenging nature of adult skeletal age estimation, its relevance and current limitations constitutes the main motivation of the research work here presented. This research aims to addresses those problems with robust yet simple computational approach, aiming to make forensic adult skeletal age estimation a more accurate, precise, and efficient process.

1.2|Hypothesis

The present work aims to lay a foundation to tackle some of the challenges of macroscopic adult skeletal age estimation, especially in its holistic or multifactorial aspect. Several authors argue in favor of multifactorial age estimation to obtain precise and accurate age estimates [48, 54, 63]. It is well known that there is not a single age indicator that can be used on its own across the entire adult age span. On the other hand, it is not necessarily true that increasing the number of age-related traits results into more accurate estimates by itself. That said, using more than one skeletal element or marker to assess age-at-death has long been pointed as fundamental to produce accurate and precise estimates. Nonetheless, multifactorial age estimation poses its own challenges and limitations, and is a topic with a clear lack of consensus [41, 49].

Conceptually multifactorial age estimation can be argued as the most effective approach for age estimation because morphological traits display different age-related trajectories, onsets, and underlying biological processes. For instance, the symphyseal face of the pubic bone 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 [64–72]. 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 to the multiple iterations of pubic symphysis aging techniques, one of the most used and favored techniques for age estimation [41], 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 [73, 74].

A multifactorial macroscopic approach to skeletal analysis does not solve, on itself, the many problems associated with determining the age-at-death. In fact, if not properly designed, this approach can become methodologically burdensome in terms of data collection and analysis. Collecting more data from the skeleton increases the likelihood of encountering redundancy, multicollinearity, and dimensionality, which impedes the straightforward interpretability and pragmatic value of macroscopic analysis. In practice, a more comprehensive analysis of age-related skeletal features necessitates a higher level of expertise in collecting the skeletal features. This is a critical issue for approaches that rely on skeleton morphoscopic analysis. Furthermore, it is common in forensic contexts for skeletal remains to be fragmentary or incomplete due to a variety of taphonomic factors, which means that not all age-related traits will be available for every unidentified deceased. From the standpoint of a practitioner, this translates into the need for computational and software tools that can fit or train age-at-death estimation models on an individual 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 with various successful applications within forensic anthropology in domains such as biological profiling or craniofacial identification [32, 52, 75–79].

1.3|Objectives

Under the premise of improving age-at-death estimation the main objective of this thesis was to propose a new method for multifactorial age estimation using an interdisciplinary approach bridging (forensic) anthropology and computer science (machine learning) through a predictive modelling framework. This major objective was decoupled into the following sub-objectives:

1) Proposed a novel macroscopic technique or protocol that better embodies a multifactorial perspective of the adult skeleton while mitigating its implicit limitations such as ease of use and scoring reliability.

2) Construct a new dataset on age-related skeletal markers. Data is primer of data-centric computational approaches and the atomic structure that enables the dialectic between anthropology and computer science.

3) Devise and validate computational approach based on machine learning technique to predict age-at-death accounting both for point and interval estimates.

4) Develop a flexible and easy-to-use software to operationalize the proposed method, accounting for specific aspects of skeletal analysis and its limitation in forensic practice.

The sub-objectives detailed above clearly elucidate the interdisciplinary nature of this work and illustrate the stages needed to establish a new biological profile estimation method. A critical view from both disciplines is needed to maximize the contribution of

each part to the whole. The development of a software that operationalizes and addresses the needs of forensic age estimation highlights the advantages and potential contributions of computationally oriented approaches in anthropology and forensics.

1.4|Document structure

The remainder of the document is structured as follows: Chapter 2 provides a historical background and a state of the art in macroscopic skeletal adult age-at-death estimation and contextualization on multifactorial age estimation; Chapter 3 presents one of the key contributions of the thesis, a novel protocol for macroscopic analysis of the skeleton and its scoring reliability analysis; Chapter 4 describes the dataset constructed for the purpose of this research, provides detailed view of the data management and processing steps, and a critical perspective of its data sources. It also provides a correlation analysis to inspect age-related variation in the skeletal traits analyzed. This chapter is heavily complemented by appendix A and B; Chapter 5 details the rationale and the mathematical formulation of the computational approach pursued in this work, and describes the machine learning approach used to construct age prediction models; Chapter 6 provides results on *in silico* validation of age prediction models constructed with the approach presented in the previous chapter and is complemented by appendix C; Chapter 7 highlights the key functionalities of the software developed throughout this research to operationalize in a practical manner the novel age-at-death estimation method proposed as the main objective of this thesis; Chapter 8 summarize the thesis conclusions, key contributions, addresses future work. The main contributions of this thesis have been peer-reviewed and published as [80–82].

2|Macroscopic skeletal age estimation: state of the art

Macroscopic skeletal age estimation of adult human remains has a long research tradition and applications in anthropology and related fields. Current chapter provides a state of the art of this topic introducing its historical background, identifying its key limitations and previous approaches to multifactorial age estimation.

2.1|Historical background

One of the first authors to thoroughly analyze the skeletal changes associated with age-at-death was T. Wingate Todd. His work can be considered the foundation of skeletal age estimation research as much of the initial research papers on adult age estimation can be attributed to or traced back to him [3]. During the 1920s, Todd extensively studied and wrote on the skeletal changes of the pubic symphysis associated with age on humans, as well as on other mammals [68, 69, 83]. He also did extensive research on the relationship between cranial sutures closure and age, accounting for confounding factors such as ancestry [84, 85]. His research provided the first methods for age estimation based both on the pubic symphysis and the cranial sutures, and for years these had been the tools available for aging skeletal remains.

Until the 1970s and the rise of anthropology from an advisory to an authoritative and fully fledged forensic science, little research work had been developed on age estimation [3]. Most of the scientific inquiry was devoted to the testing, critique, and modification of existing techniques [86, 87] — mostly a reassessment of Todd's contributions. During this time period one of the most notable exception is the work developed by McKern and Stewart [67], which introduced a components-wise method (based on Todd's phase method)

for age estimation based on pubic symphysis morphology.

The advent of forensic anthropology, during the 1970s (and its onward consolidation), brought not only the scrutiny and validation of previous age estimation methods but also new developments and methods for macroscopic skeletal age estimation. Gilbert and McKern [66] building upon previous work by McKern and Stewart introduced a component-wise method for age estimation of females based on the pubis, accounting for the sex differences observed in this skeletal indicator. The confounding effect of sex upon age estimates from the pubic symphysis became more apparent with research conducted by Suchey [88]. Hanihara and Suzuki [65] proposed new method based on the *os pubis* and using regression analysis. However, the age range (18-38 years old) of their sample limited the method applicability.

The 1980s was one of the most impactful time periods in skeletal age estimation research. Meindl and associates [89] performed an extensive validation study of Todd [68, 69], McKern and Stewart [67], Gilbert and McKern [66], and Hanihara and Suzuki [65] methods for age estimation using the Hamann-Todd identified skeletal collection. The authors concluded that Todd's method was more accurate and, to address some of the problems they identified, proposed a revision to Todd's original phase system.

Katz and Suchey [70] conducted a validation study of Todd [68, 69] and Gilbert and McKern [66] methods in large sample (n=736) of male pubic symphysis collected from cadavers of multiple ancestral backgrounds. The authors concluded that Todd's method was more accurate and easier to use and proposed a revised method by collapsing Todd's ten-phase system to a scoring system with six morphological phases. Brooks and Suchey [64] extended Katz and Suchey [70] research work including female individuals (n=237) in their sample cadaver extracted pubic symphysis. The authors research culminated in the Suchey-Brooks age estimation method, a six-phase system modified from Todd's method, which is the most well-known and used age estimation method [41]. The Suchey-Brooks

was significant contribution to forensic anthropology because it was the first pubic symphysis age estimation method created with contemporaneous reference sample making it more suitable for forensic case work.

During the 1980s, cranial suture closure as age indicator was revised by Meindl and Lovejoy [90], who proposed a new method for age estimation. Age estimation from cranial sutures was deemed as unreliable during the 1950s by Brooks [87] and Singer [86]. However, as Brooks noted the cranium is often the only skeletal element available and cranial sutures can provide a general idea of age. Meindl and Lovejoy [90] recommended the method to be applied in conjunction to other aging techniques.

The most significant contribution on age estimation research during the 1980s were the introduction of macroscopic techniques based on skeletal indicators other than the pubic symphysis and cranial sutures. İşcan and associates [91–96] developed a new method for macroscopic age estimation based on the morphological changes of the sternal end of the fourth rib. The method revealed to be very promising for forensic applications because like the Suchey-Brooks method it was elaborated using an autopsy room sample and provided good results in older individuals.

Lovejoy and colleagues [97] published a paper describing the age-related changes associated to the morphology of the auricular surface of the ilium. Based on validation studies the authors stated that their technique is as accurate as the Suchey-Brooks method. Although more difficult to apply because the metamorphosis of the skeletal indicator does not have features so conspicuous as those of the pubic symphysis. It is important to note that in a recent survey conducted by Garvin and Passalacqua [41] the Suchey-Brooks [64, 70, 88], Meindl and Lovejoy [90], Lovejoy et al. [97], and İşcan [91–96] age estimation methods remain the most used methods by forensic anthropology experts.

The process of validation and refinement of macroscopic skeletal age estimation methods, and the creation of new ones has continued throughout the 1990s to nowadays.

To handle common problems with Işcan method, correctly identify the fourth rib and the fragile nature of the sternal end, Kunos et al. [98] developed an extensive study of the first rib age-related changes. The first rib can be easily identified and recovered due to its unique morphology, and its small size make it extremely robust against taphonomic agents. Kunos et al. identified age-related modifications not only in the sternal end of the rib but also on the tubercle and the head. DeGangi et al. [99] revised the method created by Kunos et al. and based on it proposed a new one using a sample of males of Eastern European origin by applying a probabilistic method for age estimation.

Lovejoy et al. [97] auricular surface aging method was evaluated by several authors. Murray and Murray [100] found that the method was equally applicable across sexes and ancestral groups but the rate of degenerative changes on the auricular surface was highly variable. Osborne and associates [101] reached very similar conclusions regarding the effect of confounding factors upon the age-related morphology of the auricular surface of the ilium demonstrating that age-related modifications on this skeletal indicator do not depend on sex, ancestry, or secular changes. However, the authors argue that the age ranges of Lovejoy et al. [97] method (5-years intervals) are too narrow, and inappropriate for forensic purposes. They proposed a revised six-phase version of the original method.

To address a problem initially pointed out by Lovejoy and his associates [97], the difficulty of correctly identify the age-related of the auricular surface, Buckberry and Chamberlain [102] and Igarashi et al. [103] proposed new methods of age estimation using a component-wise approach. Despite the inherent limitations of the auricular surface as an age-related skeletal indicator, several authors concluded that Buckberry and Chamberlain method is superior to Lovejoy et al. technique [104–108].

The age-related morphology of the pubic symphysis continues to attract attention and the latest research trends have been focused on the quantification of age-related changes using medical imaging technology and statistical shape analysis [71, 72, 109–114].

Obtained results are encouraging, yet currently such methods present limited practical value as they require highly specialized equipment and personnel.

In terms of macroscopic analysis, Berg [115] and Hartnett [116] both revised the Suckey-Brooks method. Using modern samples, the authors proposed adding a new phase to the original six-phase system to more accurately identified age-related changes in individuals over 70 years old. Chen et al. [117, 118] proposed a new component-wise regression-based method for age estimation. This new regression-based method was compared to Suckey-Brooks method in a validation study by Fleischman [119], which concluded that Chen et al. method is more accurate for middle-age adults. Castillo and colleagues [120] proposed a novel method for age estimation based on 16 binary traits observed in the pubic symphysis and machine learning algorithms. Preliminary results show promising result for older individuals.

In recent years, several papers have made significant contributions for macroscopic adult skeletal age estimation, researchers investigated and proposed new techniques based on new skeletal indicators never analyzed systematically in adult age estimation.

Passalacqua [121] proposed a new method for age estimation of human remains based on the developmental and degenerative changes of the sacrum. Validation tests demonstrated that the method can be applied with high reproducibility, nonetheless it provides imprecise age estimates with large predictive intervals [122].

Falys and Prangle [123] studied the post epiphyseal changes occurring on the sternal end of the clavicle and verified that highly patterned degenerative skeletal changes can be observed. The authors propose a regression-based method to predict age-at-death based on three features, preliminary results demonstrate that the degenerative changes of the clavicle can be useful on the estimation of advanced age.

Several methods for age estimation based on the acetabulum have also been proposed and validated [124–140]. Literature suggests that age estimation using the

acetabulum is very promising because the acetabulum shows a stable pattern of changes with a late onset which is very useful for the elderly, it is also an anatomical region very dense and resistant.

Listi and Manheim [141] and Listi [142] have conducted statistical analysis on the utility of osteoarthrosis and entheseal changes as age-related indicators. While their results show that both skeletal conditions should not be used as the only mean of age estimation, the analysis of this indicators can be critical to identify individuals who died with more than 70 years old. Alves-Cardoso and Assis [143], Winburn [144], and Milella et al. [145] also conducted research on joint and musculoskeletal degenerative traits and their results urge to reconsider these traits as predictors of age-at-death in adults. While estimates based solely on these traits over-age individuals under 45 years old, they are important to refine age assessment in the elderly due to the late onset of this type of skeletal changes.

2.2|Limitations in skeletal age estimation

As demonstrated throughout the historical overview, there is a large set of techniques available for macroscopic skeletal age estimation. Virtually every bone of the human skeleton can be a valuable source of age-related information. Nonetheless, adult skeletal age estimation is one of the most difficult tasks an anthropologist needs to perform, and where a lack of consensus is evident. An unfortunate expression of such lack of consensus is the fact that often experts prefer or give more emphasis to a certain method without provide any statistical or scientific information to support such decision [41, 43, 49]. When multiple skeletal indicators/methods are used, final age estimates are often obtaining using simple heuristics where the anthropologist's experience often plays an important role, which leads to subjectivity being an influent source of error and bias.

2.2.1|Observation error and observer subjectivity

Intra- and interobserver error is a significant bias component in age estimation and it is always present in macroscopic techniques. Identifying age-related morphological features often relies on intricate and elaborate descriptions that may be open to multiple interpretations. A general observation in skeletal age estimation, is that the level of proficiency and familiarity of the expert with a given technique will increase the reproducibility and reliability of the age estimates [44, 47]. An elusive problem of observation error is that when observation error collides with observer subjectivity. For instance, the expert may assume that his proficiency with a given technique for certain skeletal indicator translates, immediately, to a comparable level of proficiency using another technique or methods available for the same indicator.

Several statistical tools are available to assess the reliability and reproducibility of macroscopic aging techniques. However, authors often fail to include such information when proposing a new technique. Another problem is that classical statistical indicators of reliability (i.e., Cohen's Kappa and its variants) are omnibus indices, and provide only information on the overall reproducibility of the method. Having information on which stages or phases of a given macroscopic method can be re-scored with higher reproducibility can be more informative to fine-tune and refine a method. While the quantification of observer error is an important step when developing a new method, currently no methodology is available to incorporate the effect of observation error in the final age estimate.

2.2.2|Biological variability

Unfortunately, the poor performance of adult age estimation techniques cannot be attributed solely to a lack of methodological consensus or operator error. Indeed, as

previously stated, biology is a major source of error in adult age estimation. The human skeleton's maturation and degeneration are related to chronological age (civil years passed since birth to death). However, this relationship does not appear to be a linear correlation between chronological age and skeletal morphology, particularly in late adulthood. Bone remodeling throughout life is constant, so that the age estimate depends, almost entirely, on these changes, which occur, at greater or lesser rates, in different areas of the skeleton [146]. Individual biology and its interaction with the individual's environment can explain the nonlinearity between skeletal degeneration and chronological age. While aging and growing old are universal and progressive in nature, they are highly variable at the individual level. Individual biological or physiological aging is influenced by complex interactions between the individual's genetic constitution, environment, and cultural practices [42, 147].

While there is an association between skeletal morphology and chronological age such relationship is affected by numerous factors. Such factors can accelerate or delay the maturation or degeneration process of the skeleton. Mays [17] points out that up to 60% of the variation observed in skeletal indicators used in age estimation can be attributed to sources other than age. The goal of a forensic anthropological analysis is to extract the chronological age to aid the identification process. Chronological age is a proxy variable for a set of complex factors that cannot be easily quantified, may or not be a direct product of physiological aging, but accumulate over time.

2.2.3|Skeletal reference samples

The structure of reference samples employed in the creation of age estimation method has long been seen as a source of error and bias in age estimation [50]. One problem that can be easily pointed out regarding the structure of the reference sample is its size.

Reduced sample size undermines statistical significance and support. Small samples tend to have uneven distribution of age-at-death and other factors such as sex and ancestry. Uneven distribution of age-at-death is a problem not constrained to small samples. For example, the Suchey-Brooks method [64] was based on a large sample size (n = 1225), but the age-at-death distribution is heavy-tailed towards young individuals, which is one of the reasons the method produce such poor results for older individuals. Another unrecognized issue with reference samples is their representativeness. Typically, researchers assume that their samples are representative of the general population. A documented reference sample is commonly assumed to be a representative sample; however, this assumption is frequently incorrect and unsupported. Most documented collections of human skeletal remains contain subsets of individuals that are unlikely to represent individuals of every possible age, ancestral group, socioeconomic status, or cultural group in a modern human population. A study by Komar and Grivas [148] demonstrated that even modern identified skeletal collections, carefully curated for forensic applications, do not represent the living and decedent populations from which they were drawn in terms of demographic parameters.

As reference collections are the most valuable resource in biological profile estimation research, researchers need to be aware to the bias associated to this source of information. Studies such as those of Komar and Grivas [148] do not dismiss the importance of identified skeletal collections, but force researchers to reconciled with the unbearable truth the skeletal reference collections are *manufactured* samples representative only of the general patterns of human skeletal variation, and they are the perfect example of what is called in the statistical lingo convenience samples. The structure of the reference sample is a particularly relevant issue in age estimation, because the age-at-death distribution is a critical and fundamental piece of information for age estimation methods and is linked to another issue in age estimation techniques: the computational methods used.

2.2.4|Computational and statistical methods

Computational and statistical methods employed on the creation of age estimation techniques have been a topic of debate and contention in the last years [32, 50–62]. From a computational perspective, methods such as the Suchey-Brooks technique (a phase-based method) can be viewed as a nearest-neighbor procedure. The age distribution of the reference sample is conditioned on each level of the discrete variable representing the phases of morphological metamorphosis and descriptive statistics such as the most relevant quantiles can be used to estimate age in unidentified individuals. This simple procedure can be seen as a case of multiple classical calibration [56], and from a technical point of view is unproblematic. Nevertheless, this type of procedure has the underlying assumption that the target case (the unidentified individual) is drawn from a population with same conditional distribution of age given indicator as the one observed for the reference sample.

Another technique that also operates under this assumption is linear regression. The application of linear regression in age estimation had been thoroughly discussed in Aykryod *et al.* [149], Konigsberg and Frankenberg [57] and Lucy *et al.* [56] the following remarks are worth mention: skeletal traits used in adult age estimation will rarely exhibit a linear relationship with chronological age-at-death; regressing age on the skeletal trait implies from a biological perspective that the changes in x, the random variable (skeletal indicator), cause or explain a change on y (age). From a biological point of view this is implausible, and the inverse situation occurs.

Probabilistically speaking the problem with the conditional quantiles and regression approaches is that they assume an implicit prior distribution on age-at-death, being such distribution the one observed for the reference sample. If the distribution of age-at-death from which the target case was drawn is not like the reference sample this result in bias and error on age estimates. This issue was first acknowledged among the anthropological

demography community and labeled as age-mimicry [50]. Age-mimicry becomes particularly problematic for methods built using a reference sample with an uneven or skewed age distribution. Although aforementioned methods have been criticized in the past, Samworth and Gowland [55] argue that if the assumption of the preservation of the conditional distribution of age given indicator can be supported, such methods are valuable tools with desirable proprieties such as the ease of modelling the distribution and derive point-estimates, and predictive intervals from it. In situations where the conditional distribution of age given indicator is not preserved, a solution is to assume that the conditional distribution of indicator given age is the same in reference and target sample. However, deriving age estimates is not straightforward as in the first case. A solution to extract age estimates from such distribution is via the Bayes' theorem. By using Bayes' theorem one can obtain the conditional distribution of age given indicator *f(y|x)*, from the conditional distribution of indicator given age *f(x|y)*, through the following general expression

$$f(y \mid x) = \frac{f(x \mid y) f(y)}{f(x)} = \frac{f(x \mid y) f(y)}{\int_a^b f(x \mid y) f(y) dy} \tag{2.1}$$

where *f(y)* is the prior distribution for age-at-death. The marginal density function *f(x)* ensures normalization and can be found, from the numerator, by integrating with respect to age. In practice this is normally obtained through numerical approximation. a and b are the bounding parameters for the domain of the conditional distribution, in adult age estimation we are usually interested in age between 18 to 110 years old.

The probabilistic formulation for age estimation is an approach that gained popularity in anthropology in the last years. To estimate age-at-death using this formulation is essential to know *f(x|y)* and *f(y)*. Usually *f(x|y)* must be estimated from a

reference sample, and it can be obtained with different approaches. Lucy *et al.* [56] and Martins et al. [62] propose using nonparametric density estimation with kernel methods. Boldsen et al. [58] and Konigsberg and associates [54, 150] have proposed parametric methods based on logistic and ordinal regression, where the indicator is regressed on age. The authors approach has become widely known as transition analysis. The name transition analysis derives from the fact that using this approach the slopes and intercept of the regression models of indicator on age can be converted to means and standard deviation of age-at-transition, that is, the age at which individuals move from, for example, a stage I to a stage II of the Suchey-Brooks system. The two approaches for estimating *f(x/y)* differ only in assumption made about the shape of the conditional distribution they are modelling. The first method makes no assumption on the shape of the distribution. Transition analysis assumes the conditional distribution *f(x/y)* has a shape that can be modelled with logistic, or normal (and log-normal) distribution parameterized with age-at-transition means and standard deviations.

The probabilistic approach to age estimation has several advantages over traditional statistical methods used to predict age. First, the output of equation 1 is the entire conditional distribution of age-at-death given the observed indicator. The most likely age, or the conditional expectation, can be easily extracted as it will coincide with age-value associated to the highest density of the distribution. Prediction (or credible) intervals can be obtained by computing the appropriate quantiles of the posterior distribution. The conditional (or posterior) distribution of age given indicator can be plotted, providing a visual insight on the age estimation process. Second, this approach while dependent on knowledge of *f(y)*, the age-at-death distribution of the population from which the target individual was drawn, *f(y)* is not assumed to be equal to the distribution of age-at-death in reference sample. Information on *f(y)* can be obtained from vital statistics and population demographic studies, parametric mortality models or assumed to be uniform.

This is an important propriety of probabilistic age estimation and a key-point to avoid the issue of age-mimicry.

Probabilistic methods are an important tool in age estimation, but like other techniques they also have shortcomings. To be able to compute the posterior distribution of age given indicator, *f(y|x)*, knowledge on *f(y)* is a fundamental piece of information. In forensic scenarios, such information can be hard to obtain without making strong assumptions. When age estimation is being performed in the context of paleodemography, probabilistic methods introduce a circular logic because to estimate *f(y)* for the target population we need to estimate individual age using *f(y|x)* but to do such estimate with accuracy, information on *f(y)* must be given in advance.

With no information available on *f(y)* to compute *f(y|x)*, some authors argue in favour of a uniform prior on age-at-death [57, 58]. That means that it is assumed that an individual has equal chance to die at every given age. While this is a weak prior, yet equivalent to the lack of information, it would not be a realistic model for most situation faced in age estimation or to model human mortality. When incorrectly assumed, a uniform prior can produce inaccurate age estimate and statistical artefacts such as the over-estimation of age in older individuals.

Probabilistic age estimation is also hard to generalize to the multivariate scenarios without incurring in some strong assumptions modelling the full conditional distribution of indicators given age. Conditional independence of indicators given age is often assumed to ease up computation even if such the premise is not fully address and tested with appropriate statistical tools. A recent and promising development in this approach to age estimation is the proposal of the Multivariate Cumulative Probit algorithm (MCP) by Stull and collaborators [151]. This algorithm expands and solve issues in algorithm proposed by Konigsberg [54].

Machine learning and artificial intelligence techniques such as neural networks,

random forests, and genetically evolved ensembles models have also been applied to macroscopic adult skeletal age estimation [32, 51, 53, 61, 81, 120]. Studies using such computational methods demonstrate promising results and show that this type of techniques are efficient for detecting age-related patterns from noisy morphological data. Thus, producing models that are robust against confounding effect such as sex or ancestry without imposing or making strong assumptions compared to traditional statistical or probabilistic approaches. Machine learning as a key computational tool in this thesis will be later addressed in more detail.

2.3|Multifactorial age estimation

The topic of this thesis, multifactorial skeletal age estimation, is not novel in anthropology. As previously stated, conceptually multifactorial age estimation is the most effective approach for age estimation because skeletal indicators have different age-related trajectories and underlying biological processes.

2.3.1|Contextualization

True multifactorial methods, that is, methods developed with the same reference sample and multiple skeletal indicators, are rare and typically employ no more than two or three skeletal markers [58, 61, 62, 126]. A common approach is to use several single methods independently, then combine the results based on some merging criteria. When using this method, there are two obvious issues: To begin, if the methods to be combined are not correctly constructed, which is unfortunately common, the problems and sources of error in such methods will be propagated in the multifactorial age estimate. Second, there are no guidelines for which methods should be used and how they should be combined.

Ritz-Timme *et al.* [43], Rösing et al. [46] and Cunha et al. [13] provide recommendations for which methods to use in forensic age estimation, but the problem of how to combine multiple methods remains.

Anthropologists use a variety of heuristics to combine age estimation from different methods, which frequently results in the situation where, even when the same methods are used, the final age estimate can differ greatly among experts [41]. Experts usually provide an overall age range based on the minimal or maximal age range of observed methods, the overlap of age ranges, or the average of age ranges when combining age estimates from different methods. Often, researchers provide a final age assessment based on the age range of the methods they believe are appropriate for a given skeleton, with their experience in osteological analysis serving as a guideline. While experience is important, basing a forensic age estimate on it invokes the *argumentum ad verecundiam*, or authority argument. That should not be allowed under any circumstances. A forensic method should strive to produce the same result whether applied by someone with thirty years of experience in the field or someone who is just starting out.

2.3.2|The *complex* method

Nemeskéri, Harsány and Acsádi were the first to propose a multifactorial age estimation method, the *complex* method. Their method combines the macroscopic analysis of the pubic symphysis relief and cranial sutures obliteration with the radiographic analysis of the trabecular bone pattern of the proximal humerus and femur. It was part of the recommendations that resulted from the Workshop of European Anthropologists, a symposium held in Prague in 1972 [152]. The four indicators used in this method, have a phase system that describes the progressive alterations occurring with age. The final age estimate is obtained by consulting the tables created by cross-tabulating indicators against

one another. Such tables are available in Ferembach *et al.* [152].

2.3.3|The *summary* age method

Lovejoy et al. [153] proposed a general multifactorial age estimation technique they called the summary age method. Simply put the summary age method is a method for combining age estimate obtained from different methods based on a weighted average. The way how the weights or weighting coefficients used in such averaging procedure are determined, on the other hand, is statistically sophisticated.

To compute the weighting coefficients, age estimates must be first derived by apply aging techniques available for different skeletal indicators. An inter-correlation matrix of the age estimates is computed and subject to a principal component analysis. The first principal component is assumed to represent true age. The loadings of age estimates, correlation of the age estimates for each method with the first principal component scores, are then used as weights for combining the age estimates.

The authors illustrated the procedure by combining age estimates from pubic symphysis, auricular surface of the ilium, dental wear and radiographic analysis of the femur but it should be viewed as a generalizable technique.

Martrille et al. [48] used this procedure to combine age estimates from several methods employed in validation study and concluded that this procedure provided most accurate and less biased age estimates outperforming single techniques.

2.3.4|The transition analysis method

The transition analysis framework [58, 59], more widely known for being a probabilistic approach to macroscopic age estimation, can also be seen as multifactorial age estimation technique. In this method five segments of the cranial sutures, five features

of the pubic symphysis and eight feature of auricular surface of ilium are statistically combined to compute the conditional distribution from which the final age estimate is derived. However, is important to note that the skeletal indicators are combine under a strong independence assumption. That means that using this approach, one is assuming that all morphological features are dependent on age but bare no relation to each other. While one can suppose that, conditioned on age, the cranial sutures closure has a low dependence with the changes occurring in the pubic symphysis, is difficult to sustain such argument when considering the dependence among the cranial suture segments or the feature of the pubic symphysis. Boldsen et al. [58] suggest that the lack of conditional independence will affect only the predictive intervals and not the point estimate (most likely age), but on a recent study Konigsberg [54] demonstrated that failing to correctly model the conditional dependencies in skeletal indicators can produce biased point estimates and predictive intervals.

Validation studies of the transition analysis method reached unexpected disappointing results [59, 154]. Despite the elaborated conceptual and mathematical foundation, results obtained using transition analysis are comparable to the ones obtained with less complex methods. Surprisingly, the method works quite well on older individuals which by itself is a major improvement. A noteworthy conclusion from validation studies is that combining multiple indicators improves the accuracy of age estimates. Jooste et al. [154] demonstrated that using the cranial sutures, the pubic symphysis and the auricular surface independently resulted in a mean absolute error ranging from 15.54 to 23.48 years old but combining the three indicators reduces the mean absolute error to values between 10.39 to 12.91 years depending on the chosen prior distribution. A refined version of transition analysis with expanded skeletal traits is currently being developed [155], but no published data exists on its accuracy at the moment.

2.3.5|The fuzzy logic system method

Anderson and associates [156–158] developed an algorithm for aggregating results of different methods based on fuzzy logic instead of probability theory. Their method uses the Sugeno fuzzy integral to produce what they call an age-graph from which a final age estimate is derived and visually analyzed. The algorithm requires a measure of accuracy and quality for each method to combine them. The authors demonstrate the procedure by using the correlation between the skeletal indicator and chronological age as a measure of accuracy and skeletal preservation as a measure of quality, but any other measure between 0 and 1 can be used. There is currently no published validation study for this procedure. From a computational standpoint, the method is complicated. An important step in the algorithm is the solution of an N-1 degree polynomial, which can only be done easily with numerical algorithms when N is large. The method lacks an easy-to-use implementation (i.e., software), which may explain why no one other than the authors has used it thus far. One disadvantage of the method is that it only combines age ranges obtained through other methods. If the age ranges of the methods to be combined are biased due to issues with those techniques, the bias will be propagated through the proposed fuzzy logic system.

2.3.6|The two-step procedure (TSP) method

The two-step procedure [159] is an heuristic that chronologically combines a macroscopic skeletal age estimation method, the Suchey-Brooks system [64], with a dental age estimation technique, the Lamendin's method [38]. The method is based on two simple premises: First, no single indicator can be informative and relevant for the entire human lifespan. Second, instead of a brute-force combination of indicators to provide a final age estimate, the standard approach in multifactorial age estimation, one should choose the indicator(s) that work best for a given stage of the human life span. To choose which

indicator(s) is more appropriate for a given age interval we would need to know the age of the individual and that is exactly what we are trying to estimate, the morphology of the pubic symphysis is used as a filter. In a first step, the morphology of the pubic symphysis is analyzed. If it is categorized with a stage three or less, the Suchey-Brooks system is used to provide an age estimate. That means that the individual is under 40 years old, an age range for which the Suchey-Brooks method is both accurate and precise. If the pubic symphysis shows an appearance compatible with a stage four or higher, the Suchey-Brooks method is discarded and Lamendin's regression method based on dentin translucency is applied as the second step of the procedure.

While the procedure is conceptually sound, it is obviously plagued by the problems associated to the methods it uses to arrive to a final age estimate. However, the logic subjacent to the procedure can be easily generalized and used to create multifactorial methods using other indicators.

2.4|Overview

Throughout this chapter a concise review of the state-of-art in macroscopic adult skeletal age estimation was provided. A historical analysis of this research topic clearly demonstrates a continuous refinement of old methods and the proposal of new ones. The anthropologist's toolbox has never been so rich. However, practitioners fail to reach a consensus on how to combine age estimates even when doing so is regarded as the most effective manner of estimate adult skeletal age with both accuracy and precision.

The most vexing statistical and technical problems of age estimation methods have long been identified and solutions proposed, nevertheless, most forensic anthropologist still rely on and prefer the techniques developed during the 1980s [41]. This can be explained by the fact the new computational approaches Transition Analysis [58, 59] and Sugeno

Fuzzy Integral [156–158] are mathematically complex and problem-tailored solutions that do not have an easy-to-use implementation in general-purpose statistical packages and environments used by anthropologists and osteologists. There is a need of computational resources, i.e., graphical interface-based software, that makes modern computing techniques easily available to anthropologists. Such resources should allow an easy fitting and deployment of predictive models in age estimation. Without an easy-to-use implementation, age estimation techniques based on advanced mathematical and statistical algorithms are just an interesting academic exercise without practical value for the forensic anthropologist.

3|A macroscopic method for adult skeletal age estimation

Current chapter introduces a novel macroscopic method designed for the purpose of multifactorial skeletal age estimation in adults. Method rationale, scoring systems, and reliability analysis are presented in detail.

3.1|General principles, foundations, and rationale

A principled method design is key to tackle the issues that arise in age estimation and mitigate its effects bearing in mind that no method is a perfect one-fits-all solution.

3.1.1|Skeletal traits as age-related biomarkers

Skeletal elements and features analyzed should meet certain criteria in order to be asserted as good biomarkers for age estimation [73, 160]. The criteria and conditions that define a good skeletal age marker, are according to Milner and Boldsen [73] quite simple:

(1) the features analyzed should display a strong, progressive, and monotonic relationship with chronological age-at-death.

(2) the rate and nature of change observed in the features used for age estimation should follow a consistent pattern across individuals, nonetheless, accounting for other biological parameters such as sex or biogeographic origin.

(3) the features analyzed should be possible to score (macroscopic data) or measure (metric data) in a reliable and consistent manner.

In subadults, these three criteria, are to a great extent easily fulfilled. The biomarkers analyzed in this context such as bone and teeth development exhibit a progressive relationship characterized by high linear correlations with age, can be quantified with consistency and reliability, and under normal conditions similar trajectories are observed for individuals that share the same biological parameters (i.e., sex or biogeographic affinity). Opposite to this scenario, in adult age estimation no skeletal structure seems to meet all the criteria on its own and, as stated by Kemkes-Grottenthatler [160], *all markers employed in skeletal age assessment are inherently flawed.*

How does one reconcile this inherently flawed nature of age-related markers with the need of more robust skeletal age estimation methods? A solution and answer to this conundrum is to employ a principled design when proposing new methods of age estimation. A method should strive to be comprehensive and incorporate features from as many skeletal elements as possible. More precisely, a method should encompass different aspects of skeletal development and degeneration as expressed by different skeletal elements or anatomical units. For instance, the symphyseal face of the pubic bone has been systematically studied with several methodological proposal presented in the literature, ranging from the pioneering studies that establish the morphological analysis of this skeletal marker as an age estimation technique to modern fully computational frameworks for age estimation [64–72]. However, other skeletal markers and regions that can convey important age-related information such as the degeneration of vertebral bodies and joint margins, or the roughening of muscle and tendon attachment sites have received scarce attention as aging markers. The unimpressive accuracy and precision associated to the multiple iterations of pubic symphysis aging techniques, one of the most used and favored techniques for age estimation [41], 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 [73, 74].

Following this first principle and envisioning the whole skeleton as a biomarker for age-at-death estimation, the first and third criteria from Milner and Boldsen [73] are more

likely to be fulfilled given that the skeleton analyzed as a global entity is more likely to follow a (strong) monotonic relationship with age-at-death, and the rate and nature of skeletal changes have a greater chance to be consistent across people because an holistic approach can encapsulate intra and interpersonal variation with greater finesse.

3.1.2|Challenges and issues in multifactorial approach

Using more than one skeletal element to assess age-at-death has long been pointed as fundamental to produce accurate and precise estimates, nonetheless, multifactorial age estimation poses its own challenges and limitations. A multifactorial approach to the analysis of the skeleton does not solve, on 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 running into redundancy, multicollinearity and a dimensionality that hinders the straightforward interpretability of the analysis. From a practical 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.

3.1.3|Macroscopic analysis as pattern recognition

The macroscopic analysis of human skeletal remains is the most common and elementary attribution of a forensic anthropologist, yet a crucial and distinctive skill. This type of approach is prevailing in all aspects of an osteological analysis and is particularly dominant in adult skeletal age-at-death assessment and the data generation mechanism for the most commonly used methods [41, 64, 90–92, 94, 96, 97]. The macroscopic inspection of the skeletal remains is, to the qualified expert, an inexpensive and straightforward manner to infer the age of the deceased. The visual and tactile analysis of the bone and its

morphology is both a cognitive and data generation mechanism that enables the process of age estimation. Is through this type of process that most of the anthropological data in biological profile estimation is acquired, so it is vital to have a basic understanding of the cognitive processes involved. Cognitive sciences, in particular cognitive psychology, can offer some insights to better understand macroscopic skeletal analysis and design better techniques of examination and age-at-death estimation.

Cognitive psychology aims to understand human internal psychological mechanisms or cognitive processes. Taking the mind and human brain as information processing unit as its theoretical core, it focus on the processes and mechanisms by which sensory stimuli are processed, stored, and retrieved and how it interacts with experience, existent and new knowledge [161–165].

From a cognitive point of view, the macroscopic examination of skeletal remains triggers a chain of sensory and mental processes that involve cognitive processes such as *perception*, *attention*, *memory*, and more specifically as task dominated the visual stimuli anthropological analysis evokes the cognitive process of *pattern recognition*. Human visual pattern recognition can be analyzed as a typical perception process, dependent on knowledge and experience. Cognitive scientists and psychologists define perception as the process by which humans gather information from the environment through the sensory systems and interpret such information. Pattern recognition in this context refers to the ability of recognize objects and patterns in the environment [163, 165]. A pattern recognition process or mechanism is initiated by sensory stimuli, which in macroscopic analysis of human remains is the visual and tactile inspection of the skeleton. By carefully inspecting the skeletal remains aims to detect key characteristics or patterns that are suggestive of the age of the individual.

Several theories have been proposed to explain visual perception and pattern recognition; such theories can be aggregated into two major clusters around the theories of direct perception and constructive perception. Direct perception emphasizes bottom-up processing, arguing that perception is mediated by direct acquisition of information from the environment. Constructive perception, on the hand, postulates perceptions are actively

constructed by selecting, processing, and merging the interpretation of stimuli with memory and previous knowledge using a top-down strategy. In age-at-death estimation, and other aspects of biological profile estimation that rely on macroscopic analysis, bottom-up and top-down cognitive mechanisms are actively applied to engage in feature extraction, correspondence matching and comparison with reference representations (internal or external).

How age-related information is recorded using macroscopic approaches resonates with important theories and models in cognitive science and psychology. One approach is to analyze the entire anatomical structure, defining several modal phases through which the region of interest evolves during the development and senescence process. This approach assumes that age-related morphology can be associated to certain *gestalt* where the individual parts are related as a whole (an idea that can be found in *gestalt* psychology). The modal phases are associated to detailed anatomical descriptions, illustrations, photographs, or casts. One of the most relevant examples of this approach is Brooks and Suchey [64] six-phase system for age estimation based on the pubic bone which improved upon the pivotal ten phase developed by Todd [68, 69]. Parallel to phase-based approach other authors proposed component-based systems, this approach assumes that more expressive aging pattern can be captured by analyzing and scoring independently the features that are perceived as whole in the phase-based systems. For example, instead categorizing the morphology of the pubic symphysis with a six-phase system one can develop different, ideally less complex and more objective, scoring systems for different aspects or features of this skeletal marker. Imagine a hypothetical method that scores five binary traits of the pubic symphysis, contrary to Brooks and Suchey [64] six-phase system, such method allows (theoretically) to express 32 patterns (2^5). The increased representational capability is one of the reasons why the component-based approach is appealing for age estimation, this approach is also more suited for statistical pattern recognition.

One aspect is especially important to consider regarding the cognitive basis of pattern recognition, the role of the perceiver. Several low-level mechanisms such as feature

analysis, synthesis and prototype matching are involved in the recognition of specific patterns. In humans, object and pattern recognition also involves memory which means the prior knowledge, beliefs, experience, and even prejudice are relevant factors for the accurate identification and interpretation of a pattern [162, 163, 166]. In the context of anthropological analysis, this characteristic and intrinsic aspect of human cognition has been often recognized as a major disadvantage of morphoscopic methods. Both phase and component-based methods rely on the discretization of the continuum of morphological variation and its typification according to predefined categories, either broad and global or restricted to specific features. Such process relies on detailed linguistic descriptions, exemplary casts and illustrations, all prone to subjective interpretation which translates into both intra and inter-observer inconsistencies.

3.1.4|Component-based approach

For the new scoring procedures proposed in this work a component-based approach was adopted when investigating both new and established skeletal markers for age-at-death assessment. Despite the large number of features analyzed in this proposal, all skeletal features are limited to discrete variables with no more than three levels or stages. Such specifications were established during the several iterations of the development and refinement of the system proposed, and guidelines from the literature. Shirley and Montes [167] empirically addressed the old methodological debate of phase *versus* component-based approach. Their study quantified observation error of a phase and a component-based method. 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 rationale for limiting the maximum number of coding states is quite simple, as this number increases so does the difficulty to differentiate adjacent states which results in observation errors and inconsistencies. This limited expression on the states for specific component or feature helps translate more easily, in this context, the language of the skeletal morphology into the language of the problem it aims to solve. For instance,

the binary traits represent the absence or presence of specific skeletal features but at the same time directly translate into *young* or *old* age.

3.2|Scoring procedures

The subsequent sub-sections provide detailed information about the scoring systems developed to incorporate this new macroscopic protocol to assess age-related skeletal traits. The development of new scoring systems was prompted by the necessity to standardize a data collection and generation mechanism that was more aligned with a multifactorial approach to skeletal age estimation and more suitable multivariate data analysis. Analyzing multiple traits also offsets the limitation intrinsic to specific traits when analyzed on their own [160].This was done with practical considerations such as observation error and application simplicity in mind.

3.2.1|Cranial and palatine sutures

The scoring system used for the cranial and palatine sutures consists on a modification and binarization of the proposal by Boldsen et al. [58]. This system was select because it incorporates much of the rationale of older methods for scoring ectocranial sutures (neurocranium) and the palatine sutures [84, 85, 90, 168–172]. The simplification to a binary scoring system was a result of 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 3.1 should be applied to nine segments from the palatine, the sagittal, coronal, lambdoid sutures (Table 3.2, Figure 3.1 and Figure 3.2).

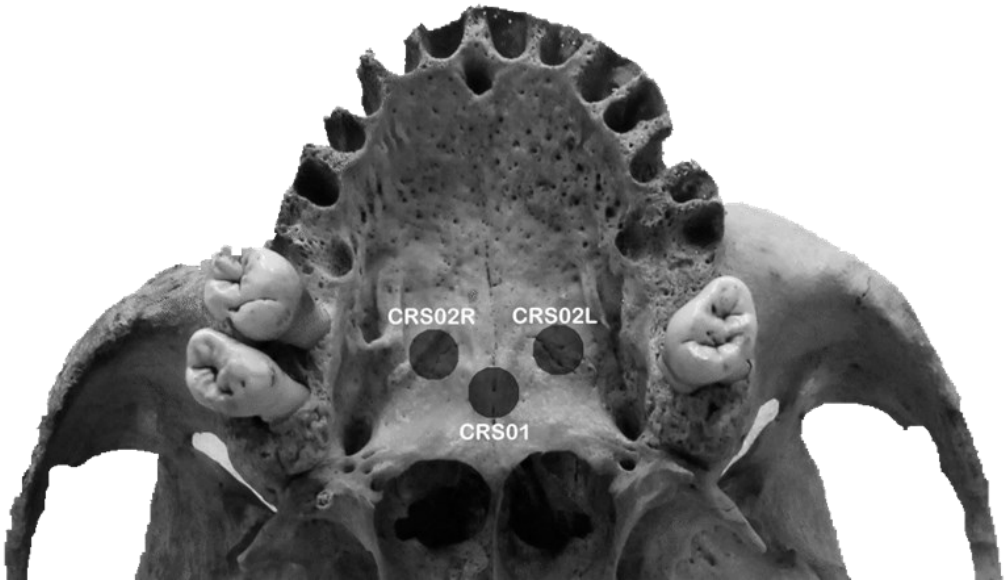

Figure 3.1 Location of palatine suture analysis segments.

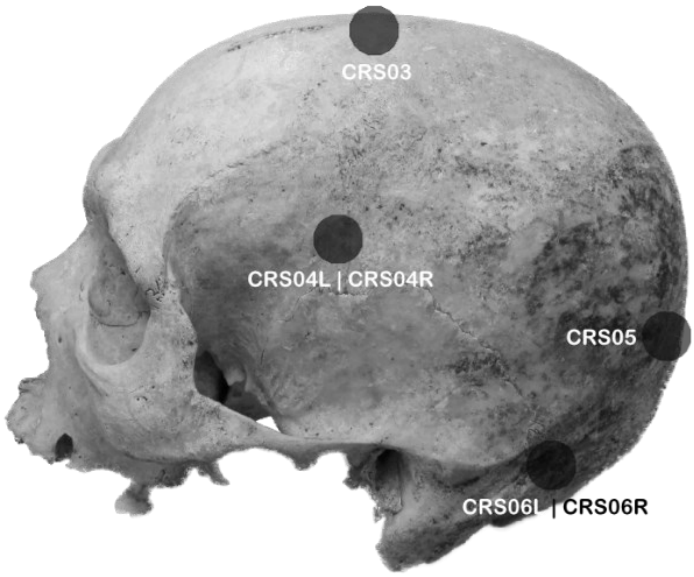

Figure 3.2 Location of cranial suture analysis segments

Table 3.1 Scoring system for suture obliteration.

Stage 0	[Open or juxtaposed] The sutural segment is characterized by a distinguishable gap between the cranial bones. The sutural gap might be narrow and the bones tightly juxtaposed.
Stage 1	[Partially obliterated or totally obliterated] The sutural segment is partially obliterated or totally obliterated. Remnants of the suture might be visible as scattered bony bridges or grooves.

Table 3.2 List of cranial and palatine suture segment analyzed.

CRS01	Palatine (posterior median)
CRS02L	Palatine (transverse, left)
CRS02R	Palatine (transverse, right)
CRS03	Coronal - Sagittal (pars bregmatica)
CRS04L	Coronal (pars pterica, left)
CRS04R	Coronal (pars pterica, right)
CRS05	Sagittal - Lambdoid (pars lambdica)
CRS06L	Lambdoid (pars asterica, left)
CRS06R	Lambdoid (pars asterica, right)

3.2.2|Vertebrae development and degeneration

The fusion of the bodies of the first and second sacral vertebrae is 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 [121, 173, 174]. This trait was assessed with a binary scale described in Table 3.3. To incorporate both metamorphic and degenerative traits of the vertebral column a three stage scoring scheme was devised building upon previous work from Snodgrass [175], Watanabe and Terazawa [176], and Alberts et al. [177]. The first two methods focus on the degeneration and osteophyte formation on the margins of the vertebral bodies while the last work focus on the development of the vertebral epiphyseal rings and body morphology. The proposed system, Table 3.4, applies to superior and inferior surfaces of the third to seven cervical vertebrae, the first to fifth lumbar vertebrae and superior surface of first sacral vertebra. Table 3.5

list all features analyzed in the axial skeletal (excluding sacral auricular surfaces). Examples depicting the traits analyzed are given in Figure 3.3, Figure 3.4 and Figure 3.5.

Table 3.3 Scoring system for S1-S2 fusion.

Stage 0	[Fusing] S1 - S2 fusion is incomplete. On the anterior surface of the sacrum, there is a gap* between the sacral bodies of the S1 and S2.
Stage 1	[Fused] The sacral bodies of the S1 and S2 are completely fused. No gap* is visible on the anterior sacral surface.

* The gap seen on the anterior surface of the sacrum should be a discontinuity between the sacral bodies extending for more 10 millimeters in length.

Figure 3.3 Fusion of the S1-S2 sacral segment (frontal view, Stage 1).

Table 3.4 Scoring system for vertebral body development and degeneration.

Stage 0	[Absence of degenerative changes] a) Incomplete or partially epiphyseal ring fusion. Residual fusion line may be observed on vertebral body. Billows or radiating grooves may also be visible perpendicular to the margin of the vertebral body margin. b) The epiphyseal ring is fully fused forming an elevated border and no degenerative change is observed on the vertebral body margin. Surface is dense and compact.
Stage 1	[Transitive stage] The vertebral margin is characterized by small segments where the edge of the margin is sharp but not necessarily lipped. The vertebral body surface is characterized by a flattened aspect. The vertebral ring has compressed appearance. Microporosities might be visible but usually have a restricted spatial distribution.
Stage 2	[Presence of degenerative changes] The vertebra is characterized by its lipped and/or porous aspect. At least one large bony projection protrudes from the body margin (approximately four millimeters or more). The surface of the vertebral body is pitted and irregular. (Vertebrae fused by lipping or ossification and calcification of the vertebral ligaments (i.e., candlewax lesions) should be score as Stage 2.)

Table 3.5 List of traits analyzed in the cervical, lumbar, and sacral vertebrae.

Cervical	C3IS	C3 body inferior surface and margin
	C4SS	C4 body superior surface and margin
	C4IS	C4 body inferior surface and margin
	C5SS	C5 body superior surface and margin
	C5IS	C5 body inferior surface and margin
	C6SS	C6 body superior surface and margin
	C6IS	C6 body inferior surface and margin
	C7SS	C7 body superior surface and margin
Lumbar	L1IS	L1 body inferior surface and margin
	L2SS	L2 body superior surface and margin
	L2IS	L2 body inferior surface and margin
	L3SS	L3 body superior surface and margin
	L3IS	L3 body inferior surface and margin
	L4SS	L4 body superior surface and margin
	L4IS	L4 body inferior surface and margin
	L5SS	L5 body superior surface and margin
Sacral	S1SS	S1 body superior surface and margin
	S1S2F	S1-S2 fusion

Figure 3.4 Lumbar vertebra superior surface (Stage 0)

Figure 3.5 Lumbar vertebra superior surface (Stage 2)

3.2.3|Joint and musculoskeletal degeneration

Osteoarthrosis and entheseal changes have been traditionally analyzed in biological anthropology and bioarcheology as markers of health, biomechanical stress, and tentative indicators of physical activity patterns. According to Milner and Boldsen [73], who advocate a more detailed analysis of this type of skeletal markers, these features as a collective contribute to an increase in accuracy and precision of age estimation. The authors base such assertion on empirical evidence from an experience-based procedure where these types of skeletal traits were extensively used. Several reasons can be pointed out on 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 complex and multifactorial etiology that hinders or masks its relationship with age-at-death.

Entheseal changes have 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 affected its relationship to 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 [143–145, 178–182].

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. In initial data collection sessions for training purposes, Buikstra and Ubelaker [183] procedure was used to assess joint degeneration. The method consists in scoring of four aspects that are involved in joint degeneration: lipping, porosity, eburnation, and subchondral exostosis.

These four features are analyzed in terms of severity but also in extent (area affected), except for bony exostosis, which increases the number of variables to be collected by joint surface to seven. Using Buikstra and Ubelaker [183] to evaluate joint degeneration of the knee – inspecting the femur, tibia, and patella—would involve collecting 21 variables both on the left and right side. Extending this analysis to other joints, would drastically increase data dimensionality. This high dimensionality poses two major problems: increased chance of collinearity, which poses computational issues, and loss of pragmatic value. Similar issues were found in the initial assessment of entheseal changes when analyzed with a protocol proposed by Henderson et al. [184]. Using this method, bony areas of muscle and tendon attachment are evaluated to assess features such as bone formation, erosion, porosity, and cavitation. A total of seven variables can be collected on each fibrocartilaginous enthesis.

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 [183] and Henderson et al. [184], 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 Table 3.6 and Table 3.7.

Table 3.6 Generic scoring system for joint degeneration traits.

Stage 0	[Absence of degenerative joint changes] Joint margin is smooth and retains normal morphology. Subchondral surface is dense and smooth.
Stage 1	[Presence of degenerative joint changes] Joint margin presents osteophytes that can range from isolated bony edges to large structures on the entire joint margin or most of it rendering its normal contour irregular. Porosities may be present both on the margin and subchondral bone surface but are less common. Bony exostosis may be present on the subchondral surface. The most extreme cases are characterized by eburnation lesions and loss of articular morphology.

Table 3.7 Generic scoring system for musculoskeletal degeneration traits.

0	Stage	[Absence of degenerative musculoskeletal changes] The surface of the attachment site contour or margin is regular and smooth.
1	Stage	[Presence of degenerative musculoskeletal changes] (One or two conditions can be present) I) The contour of the attachment site is irregular or salient. It manifests as a small bony crest or enthesophyte. II) The bone surface either presents slight irregularities in the form of a diffuse granular texture or more significant types of bone remodeling such as bony excrescences, erosions, or cavitation (large perforations).

The scoring system applies to five major anatomical complexes from the upper and lower limb: shoulder, elbow, hip, knee, and ankle (Table 3.8).

Table 3.8 List of traits used to assess joint and musculoskeletal degeneration of the limbs.

Trait		Type	Complex
SC01	Scapula (glenoid fossa)	Joint	Shoulder
HM01	Proximal humerus (head)	Joint	Shoulder
HM02	Proximal humerus (lesser tubercle)	Musculoskeletal	Shoulder
HM03	Proximal humerus (greater tubercle)	Musculoskeletal	Shoulder
HM04	Distal humerus (trochlea and capitulum)	Joint	Elbow
HM05	Distal humerus (medial epicondyle)	Musculoskeletal	Elbow
HM06	Distal humerus (lateral epicondyle)	Musculoskeletal	Elbow
UL01	Proximal ulna (articular facets)	Joint	Elbow
UL02	Proximal ulna (olecranon)	Musculoskeletal	Elbow
RD01	Proximal radius (head)	Joint	Elbow
RD02	Proximal radius (radial tuberosity)	Musculoskeletal	Elbow
OC01	Os coxa (iliac tuberosity)	Musculoskeletal	Hip
OC02	Os coxa (ischial tuberosity)	Musculoskeletal	Hip
OC03	Os coxa (acetabulum)	Joint	Hip
FM01	Proximal femur (head)	Joint	Hip
FM02	Proximal femur (trochanteric fossa)	Musculoskeletal	Hip
FM03	Proximal femur (greater trochanter)	Musculoskeletal	Hip
FM04	Proximal femur (lesser trochanter)	Musculoskeletal	Hip
FM05	Distal femur (condyles)	Joint	Knee
TB01	Proximal tibia (condyles)	Joint	Knee
PT01	Patella (articular facets)	Joint	Knee
PT02	Patella (base)	Musculoskeletal	Knee
CLN01	Patella (calcaneal tuberosity, superior)	Musculoskeletal	Ankle

To enhance the analysis of these traits are provided specific scoring descriptions for the Stage 1 of some traits (Table 3.9).

Table 3.9 Stage 1 description for joint and musculoskeletal degeneration traits.

Trait(s)	Stage 1
SC01	The key aspect is the lipping of the articular margin. Lipping (irregular bony growth) should be considered present if at least one third of the margin is affected.
HM01	The key aspect is the lipping of the articular margin. Lipping is not conspicuous as in the glenoid fossa. In early stages it takes the form of sharp elevated rim that interrupts the flow of the articular surface to para-articular region. The most severe cases form a collar or ring like structure around the humeral head.
HM04	Most common degenerative trait observed is the presence of marginal osteophytes. If present, eburnation usually affects the capitulum.
UL01	The key aspect is the lipping of the articular facets. Usually, the lipping is not so marked as in other joints. Eburnation and other type of surface remodeling are uncommon.
RD01	Marginal lipping and porosities both on the surface and margin are the most common degenerative aspects. In some cases, the bone surface seems thinned out (loss density).
OC03	Osteophytic growth of the posterior cornum is common evidence of early degenerative changes of the acetabulum. Osteophytic growth of inner margin can obliterate the acetabular fossa. Lunate surface remodeling is uncommon but in severe cases eburnation can be present. The acetabular fossa may present textural changes expressed as porosities, bony growths, and granularity.
FM01	Marginal osteophytosis and surface remodeling such as the formation of bone nodules and irregularities of the contour of fovea capitis are common. In extreme cases, osteophytic activity creates an osteophytic ring around the femoral head. In some cases, the bone surface and the margin appear thinned out (loss of density).
FM05 TB01 PT01	Surface porosity and marginal lipping are the most common degenerative traits observed. Eburnation is observed in extreme cases of joint degeneration.

Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10, and Figure 3.11 provide examples of Stage 1 on 76 years old female with a generalized degenerative pattern of the major joint and musculoskeletal complexes of the upper and lower limb.

Figure 3.6 Glenoid fossa degeneration (SC01). Stage 1, Stage 1, left side, 76 y.o, female.

Figure 3.7 Proximal humerus degeneration (HM01, HM02). Stage 1, left side, 76 y.o, female.

Figure 3.8 Distal humerus degeneration (HM04, HM05). Stage 1, right side, 76 y.o, female.

Figure 3.9 Proximal femur degeneration (FM01). Stage 1, right side, 76 y.o, female.

Figure 3.10 Distal femur degeneration (FM05). Stage 1, right side, 76 y.o, female.

Figure 3.11 Proximal tibia degeneration (TB01). Stage 1, right side, 76 y.o, female.

3.2.4|Clavicle sternal and acromial ends

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 [123, 185–187]. 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 30s. Falys and Prangle [123] were the first to propose a method to score post-epiphyseal changes of the clavicle for age estimation purposes. The authors suggest a scoring system focused on surface topography, porosity, and marginal osteophyte formation, and provide a regression model for age estimation. A new scoring scheme that integrates both developmental and degenerative changes of the sternal and acromial ends of the clavicle is proposed. Full description of the traits analyzed are available in Table 3.10.

Table 3.10 Scoring system for clavicle age-related traits.

Trait(s)		Stage	Description
CLV01	Sternal end	0	Epiphyseal union at the sternal end of the clavicle is incomplete (nonunion with or without epiphysis) or partial.
		1	Epiphyseal fusion is complete. The sternal surface has smooth to finely granular texture. The surface contour maintains a normal appearance without osteophytic irregularities. Porosities (micro or macropores) are not a characterizing feature, when they occur spatial distribution is limited (less than one-third of the surface).
		2	Sternal end is characterized by a coarsely granular texture (bone exostosis). Porosity (micro and macropores) occur in more than half of the surface. The surface contour may present an irregular profile due to osteophytic activity.
CLV02	Acromial end	0	Bone surface is smooth or finely granular.
		1	Acromial end is characterized by the presence of a coalescent porosity pattern of macropores. Bone surface appear thinned out and trabecular bone may be exposed.

Figure 3.12 and Figure 3.13 provide an example on the transitive stage (post-epiphyseal, Stage 1) of the sternal end of the clavicle and its evolution to a status characterized by degenerative skeletal features (Stage 2).

Figure 3.12 Sternal end of the clavicle – Stage 1

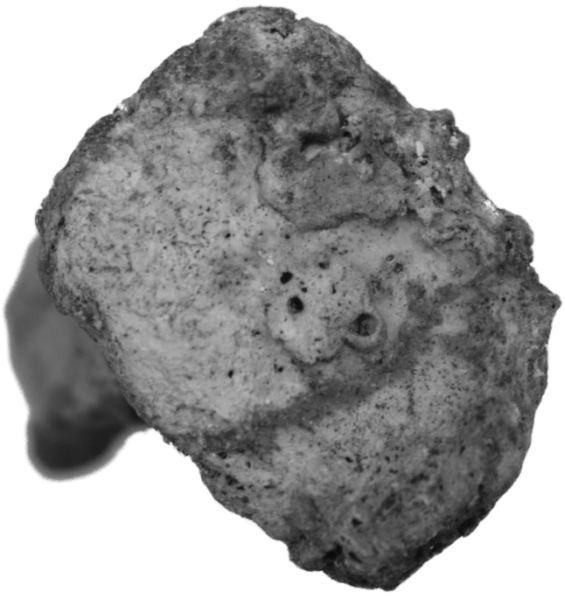

Figure 3.13 Sternal end of the clavicle – Stage 2

3.2.5|First rib costal face and tubercle

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 [91, 92, 94, 96]. Nonetheless, several disadvantages have been pointed out, such as the difficulty to identify 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. [98] described a new age estimation method based on the metamorphosis of the costal face, head and tubercle of the first rib. DiGangi et al. [99] improved upon Kunos et al. [98] work and proposed a revised method for age estimation based on the costal face and tubercle morphology. A new scoring method is proposed here that build upon previous work [98, 99]. 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 3.11).

Table 3.11 Scoring system for the first rib age-related traits.

Trait(s)		Stage	Description
RB101	Costal face	0	Costal face presents a narrow profile. The costal surface has flat profile characterized by the presence of transverse ridges or a smooth texture. The periarticular bone surface is smooth.
		1	The topography and texture of the costal face are characterized by an increasing concavity and cribriform pattern. The margins are slightly projected with scalloped edges. At the anterosuperior aspect of the margin, elongated spicules may form a rugged shaft around the costal face.
		2	Costal face is characterized by extensive ossification of costochondral interface. Ossification surrounds most of the costal face and may form a hollow shell around it. Periarticular region is rugged. Sternocostal fusion may occur.
RB102	Tubercule	0	Tubercle is characterized by rounded and smooth articular margins. The periarticular region is smooth.
		1	The tubercle facet is characterized by a coarsely granular texture. Porosities may occur in the articular surface. Lipping of the articular margins may occur. The periarticular region is rugose.

Figure 3.14 and Figure 3.15 provide an example of two opposing stages, 0 and 2, for the costal face of the first rib.

Figure 3.14 1st rib costal face – Stage 0

Figure 3.15 1st rib costal face – Stage 2

3.2.6|Pubic symphysis

The metamorphosis of pubic symphysis is the most popular osteological marker used in adult skeletal age estimation. The attention given in past to this anatomical structure is not misplaced; yet the overreliance in 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 of three features of this structure: rim development, topography, and texture of the symphyseal face. These three components (Figure 3.16) are assessed with a three-stage coding system emphasizing early metamorphic or developmental 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. Full description of the scoring system is given in Table 3.12. The proposed system is based on previous work by Todd [68, 69] and Brooks and Suchey [64].

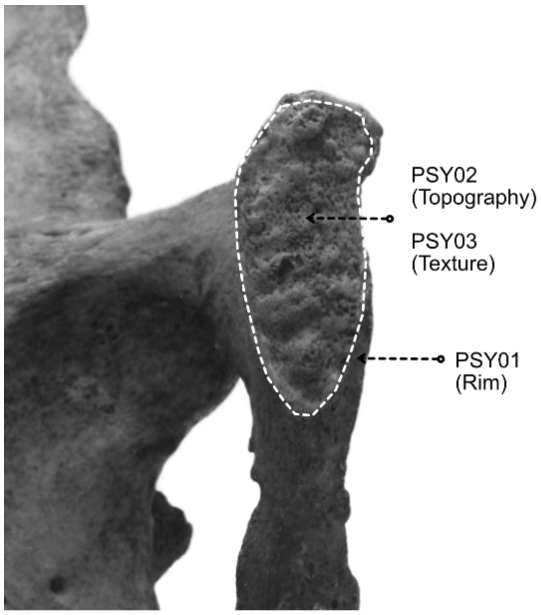

Figure 3.16 Pubic symphysis traits.

Table 3.12 Scoring system for the pubic symphysis age-related traits.

Trait(s)		Stage	Description
PSY01	Rim	0	Symphyseal rim is incomplete. In the early stage of rim formation there is a continuum between symphyseal face and adjacent structures (pubic tubercule, pubic ramus). Rim formation later evolves from an early rampart that forms on the cranial, dorsal, caudal, and ventral portions of the rim.
		1	Symphyseal rim is complete. An elevated bony rim delineates the symphyseal face demarcating it from adjacent structures such as the pubic tubercle. In some specimens, the superior segment of the ventral margin might not fully form into a rim after early formation of the ventral rampart. This condition, ventral hiatus, should not be confused with incomplete rim formation (Stage 0) or rim breakdown (Stage 2).
		2	Symphyseal rim is breaking down. Breakdown is characterized by lipping and erosion (porosity, pitting) of the ventral and dorsal margins. Breakdown of the symphyseal rim is usually associated with ligamental outgrowths and a bony plaque on the ventral and dorsal surface of the pubic bone.
PSY02	Topography	0	Symphyseal topography is characterized by a billowing pattern (alternating ridge and furrows). In early stages this pattern is very sharp but as symphyseal face flattens (Stage 1), it becomes shallow and residual (usually one patch defined by two consecutives flattened ridges).
		1	Symphyseal surface is flat and homogeneous.
		2	Symphyseal topography is irregular and depressed in relation to the symphyseal rim.
PSY03	Texture	0	Symphyseal texture is smooth to finely grained and have dense aspect.
		1	Symphyseal texture is coarsely granular yet homogeneous. Scattered porosities (micropores) may occur throughout the surface.
		2	Symphyseal texture appears eroded and is characterized by clustered porosities and irregular bony formations. Texture is less dense.

3.2.7|Sacral and iliac auricular surfaces

The description of age-related changes in the sacro-iliac joint can be traced back to Sashin [188] and Schunke [189] but its usage as an age indicator its mostly due to the work of Lovejoy and colleagues [97] and Buckberry and Chamberlain [102] on the chronological metamorphosis of the iliac auricular surface, and the age estimation method by Passalacqua [121] based on metamorphic and degenerative changes of the sacrum.

To incorporate age-related features of sacro-iliac joint, a two-component based system was developed to assess textural and marginal changes in the sacral and iliac auricular surface (Figure 3.17 and Figure 3.18). 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 feature 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 Table 3.13 and Table 3.14.

Table 3.13 Scoring system for the sacral auricular age-related traits.

Trait(s)		Stage	Description
SAS01	Texture	0	Surface is characterized by a homogeneous smooth to finely granular texture. Bone surface has a dense and compact aspect. A structured relief pattern characterized by a shallow billows or *striae* from early development stage may remain visible (residually). No porosities are observed.
		1	Surface is characterized by coarsely granular texture. Porosity occurs throughout the surface in a scattered or clustered pattern (both macro and/or micropores).
SAS02	Margin	0	The margin of the auricular surface is smooth and well defined.
		1	The contour of the auricular surface is marked by several irregularities. The margin of the surface is sharped (lipped), more commonly in the anterior and inferior apices.

Table 3.14 Scoring system for the iliac auricular age-related traits.

Trait(s)		Stage	Description
IAS01	Texture	0	Surface is characterized by a homogeneous smooth to finely granular texture. Bone surface has a dense and compact aspect. A structured relief pattern characterized by a shallow billows or striae from early development stage may remain visible (residually). No porosities are observed.
		1	Surface is transitioning from a finely granular to coarsely granular texture. Small exostoses may occur but are not a dominant textural element. Porosity (micropores) occurs throughout the surface in a scattered pattern.
		2	Surface is characterized by an irregular granular texture. Porosity is the dominant textural element—clustered distribution and presence of macropores. In overall, the surface has an irregular and eroded aspect.
IAS02	Margin	0	The margin of the auricular surface is smooth and well defined.
		1	The contour of the auricular surface is marked by several irregularities. The margin of the surface is sharped (lipped), more commonly in the anterior and inferior apices.

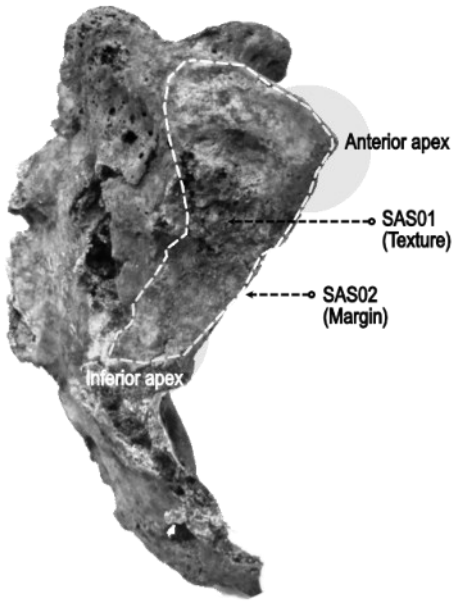

Figure 3.17 Sacral auricular surface traits.

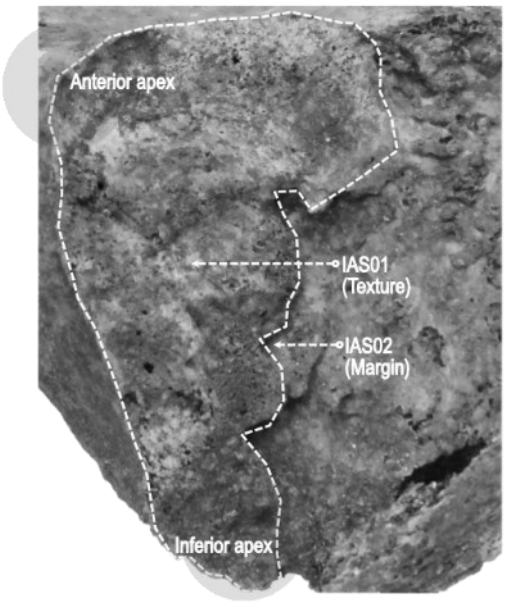

Figure 3.18 Iliac auricular surface traits.

3.2.8|Acetabulum

Several age-related changes can be documented in the acetabulum and be used for age estimation [125, 126, 128, 130, 131, 134, 138, 139, 190]. 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 proposed protocol, a three-stage scoring system for the changes occurring on the rim, posterior horn and acetabular fossa was developed (Figure 3.19). In the spirit of Calce [138] who simplified the method developed by Rissech et al [130, 131], the foundation of the scoring system presented in Table 3.16 was based on a simplification and adaptation of the method proposed by San-Millán et al. [128, 129].

The system proposed represents a significant tradeoff between of ease of scoring and representational capacity compared to San-Millán et al. [128, 129]. San-Millán et al. proposed seven traits with up to seven stages of complexity (Table 3.15). If age is to be estimated solely on the acetabulum, San-Millán et al. [128, 129] is highly recommended. The new scoring system aims only to recode and integrate acetabular morphology into more comprehensive skeletal age estimation method. The tradeoff of representational capacity is offset by ease of application and integration with other skeletal features.

Table 3.15 Correspondence between San-Millán et al. acetabular traits and new proposed system.

	San-Millán et al.	Proposed
	Groove	
	Rim shape	ACT01
	Rim porosity	
	Apex activity	ACT02
	Outer edge of the fossa	ACT03
	Texture and density of the fossa	
	Activity of the fossa	

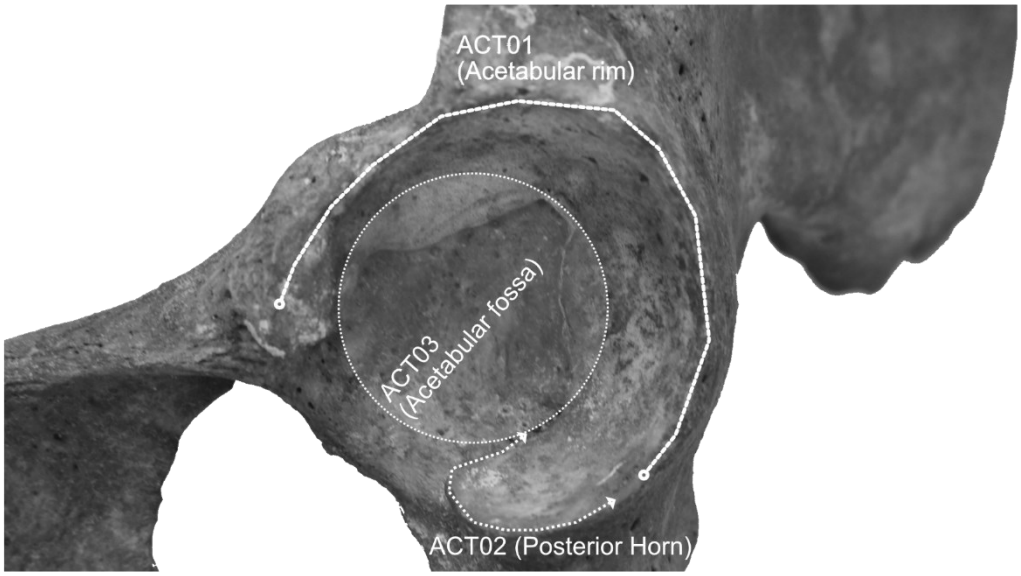

Figure 3.19 Acetabulum traits.

Table 3.16 Scoring system for the acetabular age-related traits.

Trait(s)		Stage	Description
ACT01	Rim	0	Acetabular rim is dense and smooth to the touch. The edge along the rim presents a rounded profile with no significant porosity. The area adjacent to the acetabular rim has no significant porosity and its surface is also dense and smooth.
		1	New bone formation, osteophytic activity, is visible on some regions of the rim. It is manifested as a small (approximately one millimeter) osteophytic crest along most of the rim or as a crest with a higher profile (two to four millimeters) only on a portion of the rim. The osteophytic crest is usually dense with no porosity on newly formed bone. The rim is not smooth to the touch and macroporosity may occur. Adjacent areas of the rim, such as the posterior wall of acetabulum and the region below the anterior inferior iliac spine may present porosities and textural changes that render the bone surface rough to the touch.
		2	The acetabular rim has an irregular profile as a by-product of osteophytic and osteolytic processes. A high profiled osteophytic crest (superior to four millimeters) is usually present at this stage. Such osteophytic cresting is accompanied by porotic changes on the newly formed bone which, in overall, renders the acetabulum a fragile and eroded aspect. Porosity and new bone formation can invade the lunate surface (usually below the anterior inferior iliac spine or around the ilium-ischium intersection).

Table 3.15 Scoring system for the acetabular age-related traits (Continued)

Trait(s)		Stage	Description
ACT02	Posterior horn	0	The apex is round and smooth to the touch and no bony spur is visible.
		1	The apex is rough and sharp to the touch, and a small spicule or spur can be felt (approximately two millimeters). It is circumscribed to a small part of the horn edge.
		2	A conspicuous bony spur (superior to three millimeters) is present. This proliferative feature is variable in its magnitude and extent. In extreme cases bone proliferation occupies the acetabular notch and may completely cross it or form a bony bridge. This stage usually co-occurs with more advanced stages of morphological degeneration of the acetabular rim.
ACT03	Fossa	0	The acetabular fossa center is very dense with a smooth texture. The outer edge of the acetabular fossa, along the inner border of the lunate surface, has a smooth edge with no osteophytic activity.
		1	The outer edge of acetabular fossa presents early degenerative changes. The edge is rough to the touch and the osteophytic structure can affect only a portion or the entire edge but is usually small (approximately one to three millimeters). The central region of the fossa may present a rougher bone surface (Stage 0) characterized by an increase in textural irregularities and porosities. Yet, the fossa does not have a fragile and irregular aspect due to extensive bone remodeling (Stage 2).
		2	The central region of the acetabular fossa has lost bone density and structural consistency. Porosity is a key aspect at this stage. Trabecular bone might be exposed due to extensive osteolytic and osteophytic processes. Porotic lesions have an irregular and sharp perimeter. Compared to the previous stage, osteophytic cresting, along with the outer edge of the fossa, is more pronounced both in extent and magnitude. In extreme cases, an osteophytic visor obliterates partly the fossa.

3.3|Scoring reliability

Within the scope of the approach pursued in this thesis, age estimation – and the whole biological profile estimation - can be seen as a generative process. Raw data is extracted from skeletal material, processed into (skeletal) information, which is converted into knowledge, the estimate (about skeletal age). This process, however, is not noise-free. The extent of error and noise involved in such process varies but two major components are always present, a biological and observational one.

The biological component of error in age estimation arises from the imperfect relationship between skeletal morphology and chronological age. It is particularly noticeable in latter period of the human life span where age estimates tend to be more imprecise and inaccurate. This component is irreducible and implicitly mapped within the age estimation process due to its endogenous nature: biology itself. It affects the part of the process where (skeletal) information is translated into knowledge (an age estimate). The observational (or observer) error component, on the other hand, affects the processing of raw data into information. The availability of useful skeletal information has a major dependency on the ability of the observer/investigator to extract accurate data from skeletal analysis. As most methods for skeletal age estimation are based on the encoding of morphology into discrete variables, with a very descriptive nature, observer experience and familiarity with such procedures are an important source of noise in the age estimation process. While appropriate training alleviates the effect of this component, it would be naïve to assume that it can be completely removed (even with sophisticated technological means of data).

The development and proposal of a new age estimation technique, or any other parameter of the biological profile, follows three phases according to Ferrante and Cameriere [44]: a first stage of information gathering and data collection, a second phase of computational model formulation and selection in relation to the data available, and a final phase of model validation. In current work these three stages are addressed in different moments. Chapters 5 and 6 focus on the aspects of computational approach formulation and model validation. In current chapter it will be address throughout this section the fundamental problem, yet often overlook, of data collection and observational error.

Repeatability of the proposed scoring system was evaluated through statistical analysis conducted on 50 subjects of the dataset[†], randomly selected and re-score on all possible traits. For bilateral traits only the left side were used for further intra-observer reliability analysis to avoid redundancy in reported results.

[†] Dataset constructed and used in this thesis is described on Chapter 4.

To assess observational or scoring reliability with discrete variables, such as the one described in this technique, the joint probability of the first and second observation was first computed by cross-tabulation and normalization to form a matrix A. The agreement of discrete traits, $A(X,X')$, is then given by the following mathematical expression:

$$A(X,X') = \sum_{i}^{k} a_{ii} \tag{3.1}$$

which is equivalent to the sum of the diagonal elements of A. Because many of the morphological traits of analyzed have an implicit increasing order to express a continuum of age-related progression, allocation to an adjacent morphological stage or score should not be treated as a full disagreement between observations and agreement under such circumstances is more accurately assessed with the corrected expression:

$$A(X,X') = \sum_{ij}^{k} a_{ij} w_{ij} \tag{3.2}$$

with w_{ij} being an element of matrix W, which stores the weighting factor of agreement for each element present in matrix A. Each element w_{ij} is given by

$$w_{ij} = 1 - \frac{|i-j|}{k-1} \tag{3.3}$$

where i and j are the indices of rows and columns of matrix W and k is the maximum number of stages of a morphological trait (in this proposal, $k = 2$ or $k = 3$). X' is a replicate observation of X. Chance-corrected agreement was computed using the following expression:

$$K(X,X') = \frac{A(X,X') - P(E)}{1 - P(E)} \tag{3.4}$$

where P(E) is the baseline agreement. Cohen [191–193] defined the baseline agreement as the sum of the product of the marginal distribution of both observation sessions. However, in this study, and following Navega *et al.* [136], the baseline chance of agreement was defined from a full random allocation model. This avoids situations where the marginal distribution of one observation session dominates the result of this statistical descriptor, i.e., not observing a specific score in one of the scoring sessions. To assess the statistical significance of the agreement descriptors a binomial test was performed to test if the observed agreement was greater than the baseline expected agreement. For a more detailed analysis, agreement was also computed for each specific level of each trait with the following expression

$$A(X_i, X'_i) = \frac{a_{ij} + a_{ji}}{\sum_{i=1}^{k} a_{ij} + \sum_{j=1}^{k} a_{ji}} \tag{3.5}$$

with $i=j$. The major advantage of computing detailed agreement statistics is the ability to assess the reliability of specific levels of the scoring system, which is important to understand error patterns and refine or improve a given scoring description or guideline.

In addition to the agreement descriptors Kendall's concordance coefficient, $W(X,X')$, was also computed. The computation was based on the linear relationship between this coefficient and Spearman's correlation coefficient [194]

$$W(X,X') = \frac{m\bar{r}_s - \bar{r}_s + 1}{m} \tag{3.6}$$

where $\bar{r}_s$ is the average value of Spearman's correlation coefficient among all $\binom{m}{2}$ pairs of

observers or observations sessions (m). Statistical significance was assessed based χ^2 distribution with $\chi^2 = m(n-1)W(X,X')$ and $df = n-1$, where n is the number of scored specimens. This rank-based non-parametric descriptor is not affected by the marginal distribution of X and its value is bounded between 0 (no agreement) and 1 (perfect agreement).

Scoring reliability analysis is reported on Table 3.17, Table 3.18, Table 3.19 and Table 3.20. All traits presented a statistically significant agreement and concordance between scoring data obtained by the first author in two different sessions (α=0.05). Overall average of agreement coefficient, A(X,X'), is 0.900 and concordance coefficient, W(X,X'), presents a 0.907 global average. With exception to RD01 and FM01, all traits present agreement and concordance coefficients above 0.800.

Table 3.17 Scoring reliability analysis for cranial and palatine suture traits.

					Stage				
Trait	n	A(X,X')	K(X,X')	p-value	0	1	2	W(X,X')	p-value
CRS01	41	0.878	0.756	0.000	0.848	0.898		0.874	0.002
CRS02L	42	0.929	0.857	0.000	0.957	0.800		0.891	0.002
CRS02R	42	0.929	0.857	0.000	0.957	0.800		0.891	0.002
CRS03	44	0.909	0.818	0.000	0.917	0.900		0.910	0.001
CRS04L	42	0.952	0.905	0.000	0.952	0.952		0.955	0.000
CRS04R	43	0.953	0.907	0.000	0.952	0.955		0.956	0.000
CRS05	43	0.907	0.814	0.000	0.905	0.909		0.907	0.001
CRS06L	41	0.878	0.756	0.000	0.906	0.828		0.867	0.003
CRS06R	42	0.857	0.714	0.000	0.885	0.812		0.851	0.003

The agreement and concordance coefficient values observed can be explained by the simplicity of the scoring systems used, with large number of traits binary coded. An important remark for ternary coded traits, is that mismatches only occurred with adjacent stages. An important limitation of scoring reliability conducted here is that only repeatability was assess – that is intra-observer error variation. Further inter- and intra-observer error analysis are required to assess repeatability and reproducibility among and within skeletal analysts.

Table 3.18 Scoring reliability analysis for vertebrae traits.

					Stage				
Trait	n	A(X,X')	K(X,X')	p-value	0	1	2	W(X,X')	p-value
C3IS	43	0.872	0.712	0.001	0.966	0.744	0.600	0.935	0.001
C4SS	44	0.875	0.719	0.000	0.952	0.744	0.741	0.922	0.001
C4IS	44	0.852	0.668	0.001	0.916	0.606	0.824	0.897	0.001
C5SS	43	0.884	0.738	0.000	0.938	0.706	0.865	0.919	0.001
C5IS	44	0.920	0.821	0.000	0.962	0.800	0.917	0.953	0.000
C6SS	45	0.944	0.875	0.000	0.978	0.848	0.936	0.967	0.000
C6IS	45	0.922	0.825	0.000	0.966	0.774	0.917	0.953	0.000
C7SS	43	0.953	0.895	0.000	0.989	0.875	0.930	0.975	0.000
L1IS	44	0.920	0.821	0.000	0.957	0.741	0.930	0.944	0.000
L2SS	44	0.909	0.795	0.000	0.944	0.714	0.930	0.939	0.000
L2IS	44	0.909	0.795	0.000	0.957	0.600	0.917	0.945	0.000
L3SS	46	0.902	0.780	0.000	0.941	0.690	0.923	0.939	0.000
L3IS	43	0.860	0.686	0.003	0.952	0.571	0.778	0.918	0.001
L4SS	40	0.938	0.859	0.000	0.973	0.848	0.930	0.960	0.000
L4IS	40	0.938	0.859	0.000	0.950	0.800	0.978	0.950	0.001
L5SS	46	0.924	0.829	0.000	0.962	0.696	0.941	0.958	0.000
S1SS	45	0.911	0.800	0.000	0.988	0.750	0.851	0.956	0.000
S1S2F	48	0.958	0.917	0.000	0.900	0.974		0.937	0.000

Table 3.19 Scoring reliability analysis for upper and lower limb joint and musculoskeletal traits.

					Stage				
Trait	n	A(X,X')	K(X,X')	p-value	0	1	2	W(X,X')	p-value
SC01	50	0.960	0.920	0.000	0.962	0.958		0.960	0.000
HM01	49	0.939	0.878	0.000	0.945	0.930		0.938	0.000
HM02	44	0.864	0.727	0.000	0.889	0.824		0.858	0.002
HM03	43	0.907	0.814	0.000	0.933	0.846		0.892	0.001
HM04	47	0.936	0.872	0.000	0.963	0.769		0.868	0.001
HM05	42	0.929	0.857	0.000	0.951	0.870		0.911	0.001
HM06	42	0.833	0.667	0.000	0.881	0.720		0.801	0.008
UL01	49	0.857	0.714	0.000	0.899	0.759		0.833	0.003
UL02	44	0.841	0.682	0.000	0.877	0.774		0.830	0.004
RD01	45	0.867	0.733	0.000	0.921	0.571		0.751	0.017
RD02	46	0.913	0.826	0.000	0.920	0.905		0.919	0.001
OC01	38	0.842	0.684	0.000	0.885	0.750		0.845	0.005
OC02	47	0.809	0.617	0.000	0.809	0.809		0.812	0.005
OC03	49	0.939	0.878	0.000	0.943	0.933		0.939	0.000
FM01	49	0.714	0.429	0.002	0.759	0.650		0.716	0.027
FM02	43	0.884	0.767	0.000	0.918	0.800		0.860	0.003
FM03	45	0.911	0.822	0.000	0.933	0.867		0.909	0.001
FM04	46	0.848	0.696	0.000	0.877	0.800		0.861	0.002
FM05	48	0.958	0.917	0.000	0.969	0.938		0.953	0.000
TB01	48	0.917	0.833	0.000	0.941	0.857		0.899	0.001
PT01	43	0.953	0.907	0.000	0.969	0.909		0.942	0.000
PT02	41	0.927	0.854	0.000	0.947	0.880		0.920	0.001
CLN01	43	0.907	0.814	0.000	0.929	0.867		0.900	0.001

Table 3.20 Scoring reliability analysis for skeletal age-related traits of the clavicle, 1st rib, pubic bone, sacroiliac joint, and acetabulum.

					Stage				
Trait	n	A(X,X')	K(X,X')	p-value	0	1	2	W(X,X')	p-value
CLV01	46	0.924	0.829	0.000	0.936	0.851	0.947	0.947	0.000
CLV02	41	0.854	0.707	0.000	0.880	0.812		0.867	0.003
RB101	43	0.942	0.869	0.000	0.923	0.932	0.960	0.950	0.000
RB102	42	0.905	0.810	0.000	0.923	0.875		0.901	0.001
PSY01	39	0.897	0.769	0.000	0.952	0.857	0.857	0.922	0.001
PSY02	39	0.885	0.740	0.000	0.909	0.842	0.865	0.923	0.001
PSY03	40	0.912	0.803	0.000	0.976	0.837	0.903	0.923	0.001
IAS01	44	0.932	0.847	0.000	0.949	0.800	0.960	0.960	0.000
IAS02	49	0.878	0.755	0.000	0.870	0.885		0.879	0.001
SAS01	45	0.889	0.778	0.000	0.921	0.815		0.868	0.002
SAS02	44	0.886	0.773	0.000	0.894	0.878		0.886	0.001
ACT01	48	0.917	0.812	0.000	0.958	0.778	0.909	0.953	0.000
ACT02	47	0.904	0.785	0.000	0.919	0.640	0.958	0.946	0.000
ACT03	37	0.919	0.818	0.000	0.960	0.800	0.903	0.943	0.001

The proposed method strives to be comprehensive and incorporate traits 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 an holistic approach can encapsulate intra and interpersonal variation with greater finesse [73].

4|A new dataset on age-related skeletal morphology

Data, and its availability, is of utmost importance in forensic anthropology and assumes a more central role in the approach pursued in this work. Regarding this subject Steadman [195] stated *"data is power"*, such bold assertion is particularly true in computationally heavy approaches. Data is the primer for computational thinking and reasoning, which can significantly improve forensic anthropology, resulting in more reliable and accurate techniques. Computational methods offer several benefits to the study of anthropological data. They enable more extensive and automated computation, thereby overcoming the limitations of cognitive ability and establishing stronger scientific foundations for forensic casework methods [196].

The objectives of this chapter are to provide an overview of the issue of reference data in anthropology, and to present the dataset constructed for the purpose of this research work and topic. The importance of data is frequently neglected, and its acquisition, processing and management is frequently considered a trivial aspect implicit to research. Although research work often translates into inferences and generalizations one should be bear in mind that results and conclusions from a study are always bounded to the data analyzed. The dataset here presented is the foundation of the novel age-at-death estimation technique and software proposed in this doctoral thesis.

4.1|Reference data and age-at-death estimation

The nature and structure of the reference data processed and used in anthropological studies and methodologies constitutes a crucial but often undervalued issue [197]. Any entry-level textbook on statistical analysis will acknowledge that a sample should ideally be randomly selected from representative of the population of interest. The sample size must also be sufficient for its intended purpose, such as making inferences

about the general population or building predictive models. However, this is frequently easier said than done, and forensic anthropology has long held this to be especially true [198]. In general context of scientific research, a power analysis can be used to determine the required sample size if the topic of interest has datasets that are readily accessible. Unfortunately, public open-access datasets are uncommon in anthropology which can explained by a slow adoption of an open science policy in the field, the nature of the data itself, or by a skulking Tolkien character spirit that inhabits the researchers who view their datasets as precious treasures not to be shared. Typically, forensic anthropology researchers utilize the specimens they have access to. In such circumstances, common sense has historically been the most important factor in sampling strategies which are characterized by their convenience-based nature [199].

4.2|Data source and sampling

In the context of methodological research and development of biological profile estimation techniques, identified skeletal collections also known as reference collections play a pivotal role as a data source. The dataset created for the purpose of the present work was constructed using two identified skeletal collection held at the University of Coimbra as data source: the Coimbra Identified Skeletal Collection (CISC) and the 21st Century Identified Skeletal Collections (XXIISC). A brief overview of these collections is provided in the following sub-sections.

4.2.1|Coimbra identified skeletal collection

The Coimbra identified skeletal collection is constituted of 505 complete skeletons whose acquisition is due in its majority to Professor Eusébio Tamagnini (1880-1972), director of the Anthropology Museum from 1907 and 1950. All specimens were exhumed from Coimbra main cemetery (Cemitério da Conchada) between 1915 and 1942,

representing individuals born between 1817 to 1924 and who died between 1904 to 1938. Seven individuals were previously dissected at the medical school before inhumation. With exception to nine individuals, all were of Portuguese nationality. Age-at-death ranges from 7 to 95 years old, and sex distribution is homogeneous (266 males and 239 females). Forty-five individuals are juveniles with age-at-death between 7 to 19 years old (18 males and 27 females). More detailed on this collection can be found in Cunha and Wasterlain [200].

4.2.2|21st century identified skeletal collection

The collection is composed of 300 specimens consisting mostly of elderly individuals. Individuals died between 1982 and 2012 with an average age-at-death is 81.19 for females (n=161) and 73.20 for males (n=139). Age-at-death ranges from 25 to 101 years old. This collection comprises skeletons unclaimed by the deceased relatives, it results from a protocol between the University of Coimbra and the City Council of Santarém. Portuguese law allows exhumations three years after burial to transfer skeletal remains to an ossuary, another cemetery, or a recently acquired private grave to reclaim public burial grounds. Formal notification is given to relatives. If they do not attend the cemetery or express their wishes, the skeletal remains are cremated or buried in a communal grave. Academic institutions sometimes curate donated remains, which is the case this collection. See Ferreira *et al.* [201, 202] for more details about its constitution, curation and research lines pursued with this collection.

4.2.3|Demographic structure

The dataset created through this research encompasses a total of 99 morphological features covering all key traditional age-related and other underexplored skeletal traits. Accounting for laterality it translates into 64 unique traits from the axial and appendicular skeleton collected using the new macroscopic scoring method whose rationale and details are described and explored in Chapter 3. Data was acquired on 500 individuals sampled

from the two identified skeletal collections previously described. Data was collection was pursued without access to the demographic parameters of the individuals (i.e., age and sex) which were reconciled later with the morphological data. All sampled individuals presented fully developed long bones, a screening and inclusion criterion used to assess skeletal adulthood. No individual was excluded due to pathology or taphonomy, affected skeletal structures were deemed not scorable resulting in a missing value for such trait(s). Initially, data was collected on 512 individuals but 12 were later removed due to high prevalence of missing values (more than 90%).

Table 4.1 Demographic characterization of reference data sampled from the CISC and XXI-ISC collections.

		CISC		XXIISC		Pooled collections		Pooled
		Female	Male	Female	Male	Female	Male	
	n	168	166	82	84	250	250	500
Age-at-Death (AGE)	mean(AGE)	48.48	45.33	81.84	74.88	59.42	55.26	57.34
	sd(AGE)	19.48	18.17	12.89	15.08	23.56	22.14	22.93
	min(AGE)	19	19	38	25	19	19	19
	max(AGE)	95	96	101	96	101	96	101
Year of Birth (YOB)	mean(YOB)	1877.29	1879.99	1923.87	1930.56	1892.56	1896.98	1894.77
	sd(YOB)	21.25	19.95	13.14	14.42	28.97	30.10	29.59
	min(YOB)	1830	1836	1904	1908	1830	1836	1830
	max(YOB)	1911	1917	1970	1982	1970	1982	1982
Year of Death (YOD)	mean(YOD)	1925.77	1925.33	2005.71	2005.44	1951.99	1952.24	1952.12
	sd(YOD)	6.60	7.34	3.71	3.92	38.05	38.45	38.21
	min(YOD)	1910	1910	2000	1995	1910	1910	1910
	max(YOD)	1936	1936	2012	2011	2012	2011	2012

Table 4.1 and Figure 4.1 provide a statistical and visual description of the demographic structure of the dataset considering both the collection from which the data was sourced and the sex of the individuals, for age-at-death. For this research, along with morphological data only sex, age-at-death, and year of death and year of birth were collected as demographic data. Year of birth was computed subtracting age-at-death from year of death.

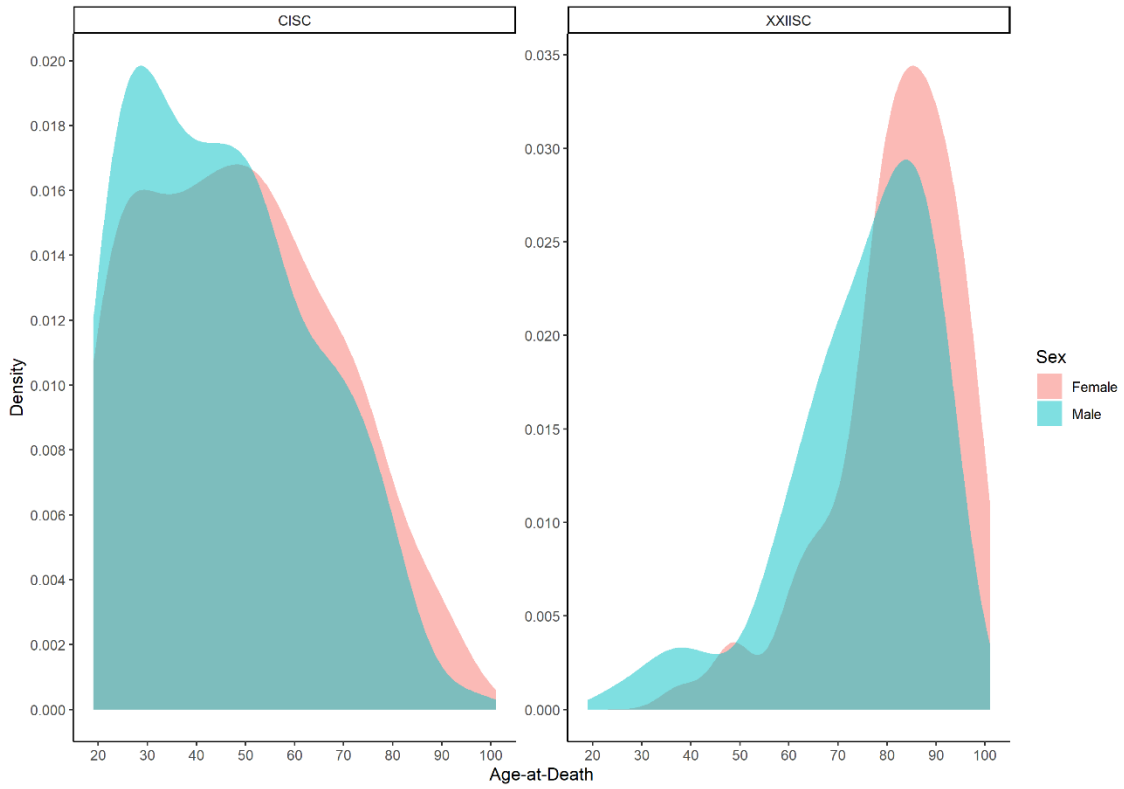

Figure 4.1 Age-at-death profiles of sampled data by sex and collection (KDE method).

4.3|Data management and processing

Multifactorial age estimation poses itself several methodological challenges, mostly related data management, processing, and analysis. A more comprehensive and expanded array of age-related skeletal traits increases the chance of running into analytical problems such as missing values, data redundancy and multicollinearity, and a feature dimensionality that hinders a straightforward data analysis, modelling, and interpretability.

The data management and processing strategy employed in this work aimed to tackle three common data-related problems, usually intertwined, faced anthropological data analysis: missing values, redundancy, and volume.

Missing values in forensic anthropology results primarily from intrinsic and extrinsic taphonomic factors. Characteristics such bone architecture and density or the ecological conditions of cadaveric decomposition and deposition play an important role on the preservation and representation of human skeletal remains [203].

Data redundancy in anthropology emerges from the body symmetry type that characterize human anatomy. Most traits analyzed in this work represent bilateral or paired data characterizing the same entity or trait across the left and right side of paired skeletal elements. The human body is not fully symmetric, yet it is not expected that the left and right diverge drastically under normal conditions. Handling bilateral is crucial minimizing redundancy and reducing dimensionality.

Data volume and quality are highly relevant for robust exploratory and inferential statistical analysis and assume a pivotal role in predictive modelling and machine learning approaches. It has already been noted that anthropological data is usually characterized by the convenience-based sampling which is limited or conditioned by factors such as availability or access to identified skeletal collection, reference series or other relevant data sources. Effective data volume, that is sample size, is also conditioned by missing values and data partitioning co-factors such sex, biogeographic ancestry or any other subsetting criteria.

The primary goal of data management and processing in this work was to cope with the challenges posed by multifactorial age estimation from a data-centric point of view. The final dataset used in predictive modelling for age-at-death estimation, the key objective of this work, is the result of several operations involving statistical inference, domain heuristics and data imputation procedures described in the sub-sections below.

4.3.1|Pooling

To maximize data volume (sample size) and age-related variation, data was pooled across the two major co-factors, documented skeletal collection and sex. Pooling data from different skeletal collections allowed for a homogenous and uniform age-at-death

distribution with a balanced representation of young and older individuals. This is a simple strategy to cope with the problem of age-mimicry and to guarantee that targeted age span is fully represented in first place [197, 204].

This first pooling operation (Figure 4.2) joined 250 male and 250 female who died at the age 19 to 101 years old (mean = 57.32, SD = 22.93) that were born between 1830 and 1982 and died between 1910 and 2012.

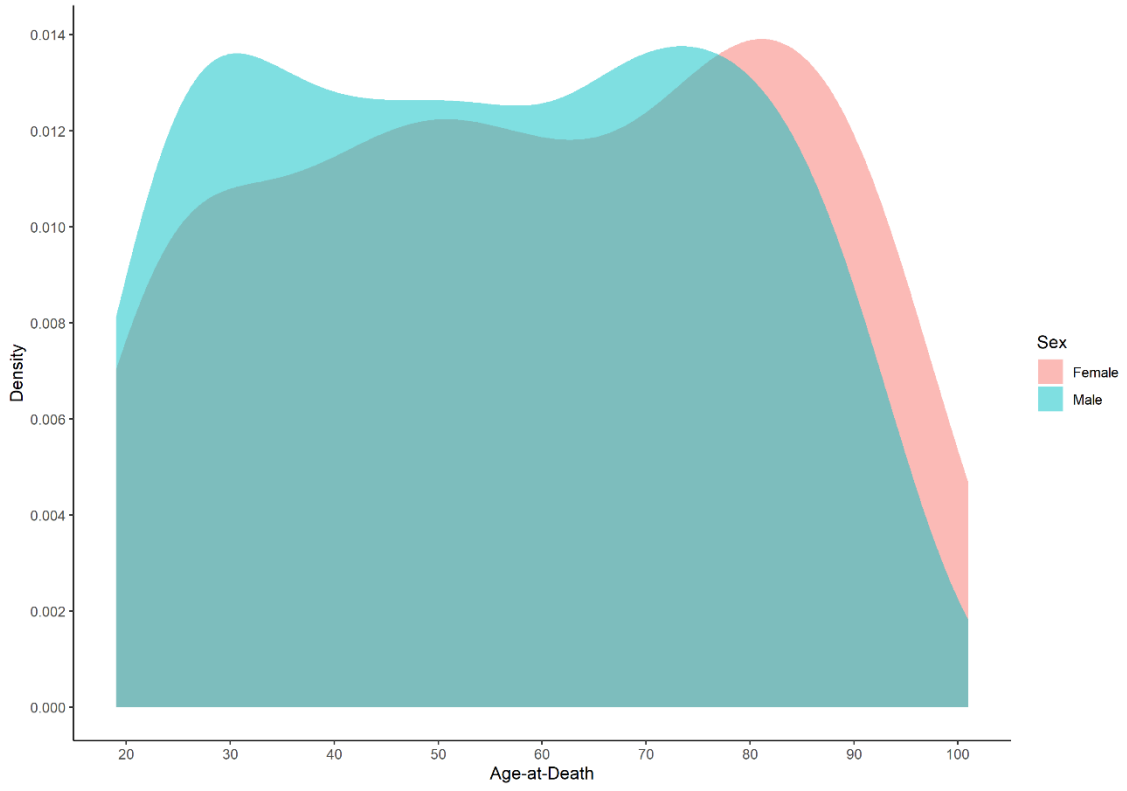

Figure 4.2 Age-at-death profiles for pooled collections by sex (KDE method).

Sexes were also pooled (Figure 4.3), while this may seem a more arbitrary choice is important to note that in forensic anthropology sex is usually estimated during casework. Pooled data models balance out the potential and pitfalls of sex-specific models and its misspecifications.

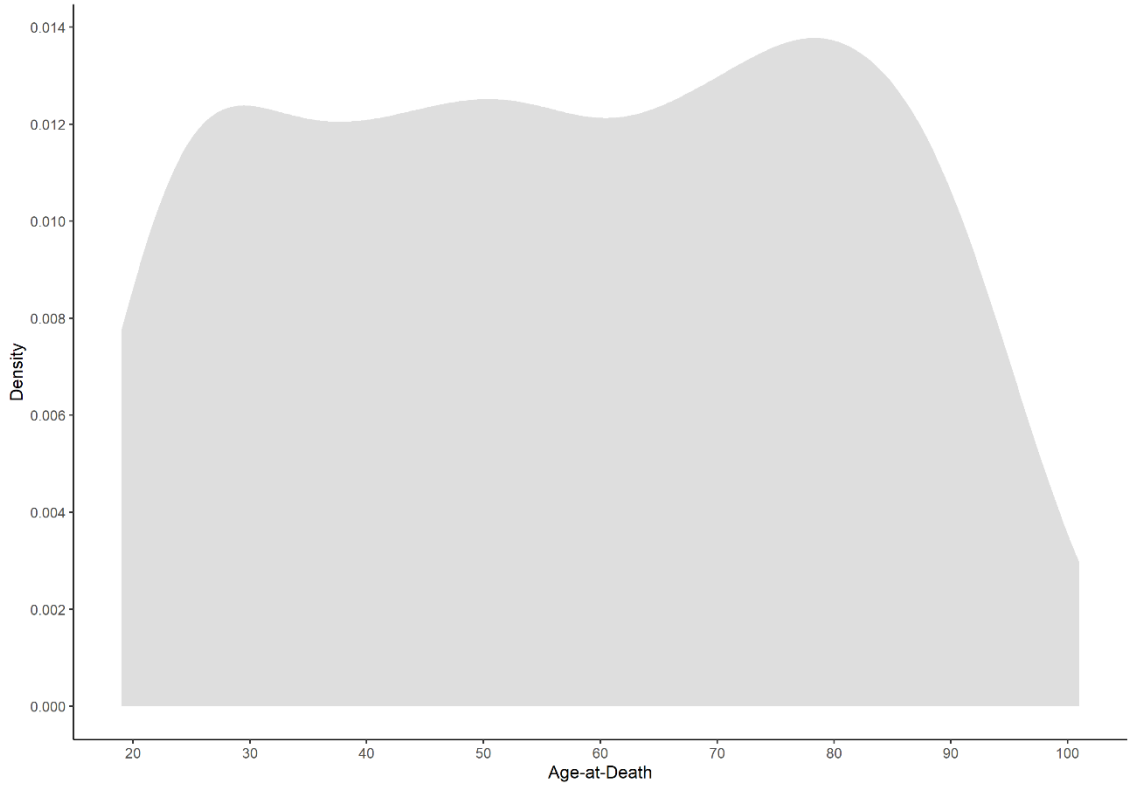

Figure 4.3 Age-at-death distribution for pooled data (KDE method).

Despite the large temporal frame represented, there is a continuum and a wide range over the age-at-death distribution that makes this dataset particularly suited for age-related research and predictive modelling.

4.3.2|Bilateral merging

To tackle the issue of data redundancy and increased dimensionality caused by paired data a simple domain heuristic was used: 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. This is consistent with anthropological data collection standards and lessens the effect of handedness and potential impact of daily or occupational activities for some traits, especially in the upper limb [183, 205, 206]. The followed step minimized redundancy and dimensionality by reducing the number of skeletal traits from 99 to 64. This procedure was applied to all paired skeletal traits with exception of the cranial and palatine sutures on which paired data exists due the quantization and segmentation into smaller parts for scoring purposes. This strategy partially deals with missing value which is address in more detail below.

4.3.3|Imputation

Due to the fragile nature of osseous material, missing data constitutes an inherent problem in skeletal (data) analysis; these issues are exacerbated in forensic anthropology, as remains in forensic contexts are frequently subjected to perimortem trauma and taphonomic factors that damages or alter key features of skeletal morphology. The presence of missing data or values poses frequently an important technical and analytical issue because certain computational and anthropological methods do not accommodate missing values or incomplete data. A common method for avoiding missing data is to exclude any case in which at least one variable is missing, also known as listwise deletion. In other cases, variables with a high percentage of missing data may be excluded from an analysis entirely. Thus, missing data can significantly reduce data volume or limit the number of variables used in a given analysis, potentially reducing the reliability of some aspects of the biological profile estimation [207].

Missing values can be handled in a variety of ways but simply ignoring them via deletion methods, such as listwise or pairwise deletion, is often an inappropriate choice which can result in biased estimates and results, a decrease of statistical power, and more important leads a loss of effective data volume. In most situations, the replacement of missing values with plausible values derived from the observation of a dataset through imputation techniques is a far superior and more valuable solution. Missing value imputation strategies, that is the reconstruction of missing values with plausible values, are underexplored in anthropology, particularly in forensic anthropological research. However, Howells [208] almost 50 years ago suggested three strategies to cope with missing values: 1) substitute missing value with the mean of the target variable; 2) apply regression models to infer missing values from observed data; and 3) to make a "careful guess". Howells's proposed strategies emerged in the context of multivariate analysis of cranial data but can by expanded to other domains of anthropological data processing and analysis, and since his time advances in computing and data analysis have permitted novel algorithms for data imputation in biological profile estimation [75, 207].

In this work a nearest neighbor procedure combined with heuristic for dealing with bilateral data was used to impute missing values. As previously described the left side for bilateral traits was used as the main source of data, and when missing the right side was use 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 its nearest neighbor. Jaccard similarity on one-hot encoded data was used to compute the nearest matches. The most notable characteristic of this strategy is that imputed values are occurring values observed in similar skeletal patterns and not estimates created by a model. Due to its nonparametric nature nearest neighbor imputation avoid model misspecification and operates under a simple yet vital feature in predictive modeling, and implicit assumption in age-at-death estimation: similar patterns produce similar estimates or predictions.

A simple nearest neighbor ($k = 1$) according to Beretta & Santianello [209] 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. This finding guided the strategy selected, which favors simplicity and minimal data distortion over imputation accuracy that is the dominant feature in more complex techniques such a random forest based algorithms for data imputation [210–212]. Imputation do not to aim "create" data where it is missing but avoid data loss, which is inevitable in forensic anthropology and exacerbated with simple deletion methods.

Missing values represented 9.52% of the total entries of the dataset when bilateral data was considered and 6.89% when the domain heuristic described was first applied as an imputation mechanism and strategy to handle bilateral data redundancy. A detail missing value percentage by trait analysis is given in Appendix A.

4.3.4|Inference

Data processing via data pooling, bilateral merging and imputation was essential to establish a new dataset for multivariate analysis and predictive modelling with a machine learning approach in the context of age-at-death estimation. This important phase of data management was not performed in an arbitrary fashion but grounded by statistical inference and guided by pragmatical aspect that should be account in forensic casework. Statistical description and inference were performed on unprocessed data and is available in detail in tabular format in Appendix A.

The association of skeletal morphology and sex was assessed through Cramér's V statistic [213]. Potential sex-related difference of conditional age-at-death distribution on trait stages were evaluated with Kolmogorv-Smirnov statistic [214, 215]. Only 10.1% (10/99) of the skeletal traits marginal distribution presented statistically significant differences among sexes. Overall, sex explains less than 4% of the variation observed in marginal trait distribution. Palatine sutures present the highest amount of sex-related variation, explaining approximately 27% of the variation in marginal distributions. Sex-related differences of conditional age-at-death distribution on trait stages were detected for 7.07% of the analyzed traits. For C4SS, HM02, OC01, FM04 and ACT02 it only affected

the age-at-death distribution for maximal trait expression (Stage 1 or 2). Only for CR01 and CR03 sex-related differences of conditional age-at-death distributions were detected on both stages.

Bilateral asymmetry analysis was conducted by computing the proportion of mismatch between paired and statistical significance assessed using Bhapkar [216] marginal homogeneity test. Laterality-related differences of conditional age-at-death distribution on trait stages were evaluated with Kolmogorv-Smirnov statistic [214, 215]. Overall, data presents an average mismatch between left and right side of 4.70%, with males and females presenting an average value of 5.00% and 4.40% respectively. Not considering sex, 6 of 39 bilateral traits presented a systematic and statistically significant discrepancy between sides. Humeral traits (HM02, HM03, HM04, HM06) presented a systematic discrepancy favoring the right side (7.4 % bias on average). OC03 and IAS01 presented discrepancy favoring the left and right respectively but low on magnitude (3.6% and 3.4% mismatches between sides). Sex-specific analysis revealed a similar pattern, with upper limb traits presenting discrepancies proportions biased toward the right side (HM02 and HM04 for males, HM06 and UL01 for females, and HM03 for both sexes). Laterality had no impact for conditional age-at-death distribution on trait stages, both for pooled and sex-specific analysis.

4.4|Age-related skeletal variation

In Chapter 3 were mentioned the criteria that define a good age-related skeletal marker, among which having a strong, progressive, and monotonic relationship to chronological age topped as one of foremost importance. To predict or estimate age-at-death from skeletal morphology implies both from a biological and computational perspective that some type and degree of association or statistical signal exists. To assess age-related signal of the skeletal traits analyzed in this work a univariate and multivariate statistical analysis using a correlation-based and hierarchical clustering approach.

Univariate correlation analysis was based on Spearman's rank correlation coefficient (ρ) and Pearson's correlation ratio (η^2). These two effect size statistics offer a nonparametric, nonlinear and interpretable way to measure the degree of association between age-at-death and skeletal morphology. Spearman's rank correlation coefficient is computed in the general case as

$$\rho = \frac{\mathrm{cov}(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{4.1}$$

where $\mathrm{cov}(R(X), R(Y))$ is the covariance of the ranks of X and Y converted with $R(X)$ and $R(Y)$ respectively. $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ are the standard deviations of the rank variables. This coefficient is equivalent to compute Pearson's linear correlation coefficient on the ranks of the data. Spearman's coefficient evaluates how well the relationship between two random variables can be explained by a monotonic function. Monotonicity is an important characteristic for age-related markers in adults because it guarantees progressive skeletal changes with age-at-death, even if not explained by a linear effect.

Pearson's correlation ratio (η^2) measure the proportion of variation in a numeric random variable that is explained by the grouping effect of a discrete random variable. It is computed as

$$\eta^2 = \frac{\sum_x n_x(\bar{y}_x - \bar{y})^2}{\sum_{x,i}(y_{xi} - \bar{y})^2} \tag{4.2}$$

where $\bar{y}_x$ is the mean of Y for category x of X and $\bar{y}$ is the mean of Y. It can also be written as the ratio of the sum of the weighted variance of Y across categories of X by the total variance of Y, hence its name. It is bounded between 0 and 1, with 0 indicating no dispersion among different categories and 1 indicating no dispersion within the respective categories or stages. This statistical descriptor has an interesting feature in that it can be used as an omnibus index, but it can also be computed for each category or stage

individually, providing a more detailed analysis of the relationship between each skeletal trait and age-at-death. Statistical significance can be obtained from a χ^2 distribution with $\chi^2 = \eta^2 n$ where n is the sample size, and k-1 degrees of freedom (k is the number of categories) [217].

Table 4.2, Table 4.3 and Table 4.4 condense the statistical descriptors for this analysis. Appendix B complements this analysis with a descriptive analysis of unprocessed data both by sex and laterality. All skeletal traits included in this protocol show a statistically significant relationship with age-at-death, although with variable effect size.

Table 4.2 Correlation analysis of skeletal traits with age-at-death for joint and musculoskeletal degeneration traits.

						Stage η^2		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
SC01	0.784	0.619	309.634	0.000	500	0.721	0.519	
HM01	0.708	0.505	252.342	0.000	500	0.467	0.551	
HM02	0.707	0.502	250.843	0.000	500	0.486	0.519	
HM03	0.589	0.346	173.000	0.000	500	0.229	0.522	
HM04	0.432	0.186	93.231	0.000	500	0.048	0.693	
HM05	0.663	0.439	219.314	0.000	500	0.249	0.737	
HM06	0.743	0.552	275.919	0.000	500	0.396	0.754	
UL01	0.577	0.335	167.338	0.000	500	0.183	0.620	
UL02	0.562	0.318	159.056	0.000	500	0.206	0.508	
RD01	0.380	0.145	72.325	0.000	500	0.032	0.636	
RD02	0.709	0.505	252.592	0.000	500	0.572	0.437	
OC01	0.709	0.506	252.851	0.000	500	0.498	0.514	
OC02	0.711	0.511	255.429	0.000	500	0.634	0.394	
OC03	0.789	0.628	313.817	0.000	500	0.685	0.569	
FM01	0.579	0.337	168.606	0.000	500	0.173	0.660	
FM02	0.515	0.267	133.251	0.000	500	0.105	0.682	
FM03	0.764	0.585	292.267	0.000	500	0.529	0.646	
FM04	0.754	0.571	285.352	0.000	500	0.585	0.556	
FM05	0.658	0.435	217.725	0.000	500	0.318	0.604	
TB01	0.559	0.312	156.146	0.000	500	0.172	0.588	
PT01	0.611	0.374	186.964	0.000	500	0.231	0.619	
PT02	0.727	0.531	265.748	0.000	500	0.475	0.598	
CLN01	0.758	0.579	289.306	0.000	500	0.581	0.576	

Table 4.3 Correlation analysis of skeletal traits with age-at-death for standard skeletal markers (clavicle, 1st rib, pubic symphysis, sacroiliac joint and acetabulum).

						Stage η^2		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
CLV01	0.851	0.729	364.626	0.000	500	0.993	0.613	0.633
CLV02	0.710	0.507	253.428	0.000	500	0.590	0.424	
RB101	0.763	0.590	294.834	0.000	500	0.975	0.038	0.787
RB102	0.776	0.607	303.462	0.000	500	0.773	0.453	
PSY01	0.711	0.523	261.291	0.000	500	0.968	0.001	0.731
PSY02	0.731	0.549	274.351	0.000	500	0.912	0.017	0.791
PSY03	0.718	0.536	267.996	0.000	500	0.946	0.093	0.550
IAS01	0.789	0.631	315.435	0.000	500	0.902	0.150	0.561
IAS02	0.731	0.539	269.408	0.000	500	0.673	0.413	
SAS01	0.632	0.398	199.074	0.000	500	0.228	0.671	
SAS02	0.704	0.499	249.481	0.000	500	0.542	0.453	
ACT01	0.782	0.625	312.327	0.000	500	0.663	0.362	0.825
ACT02	0.818	0.674	337.107	0.000	500	0.865	0.043	0.671
ACT03	0.804	0.662	330.860	0.000	500	0.829	0.094	0.741

Global Spearman's correlation coefficient (ρ) average value is 0.683. This statistic assumes its minimum average value for the cranial and palatine suture traits, 0.399 (0.297 to 0.518). Vertebral body traits show an average value of 0.821 (0.794 and 0.845), upper limb traits an average value of 0.623 (0.380 to 0.784), lower limb traits an average value of 0.678 (0.515 to 0.789), clavicular traits an average value of 0.780 (0.710 to 0.851), 1st rib traits an average value of 0.769 (0.763 to 0.776), pubic symphysis traits an average value of 0.720 (0.711 to 0.731), sacroiliac joint traits an average value of 0.714 (0.632 to 0.789), and the acetabular traits an average value of 0.801 (0.782 to 0.818).

Correlation coefficients provide insight into the ability of skeletal traits to predict age-at-death while making few assumptions about the underlying data [15]. Pearson's correlation ratio (η^2), is particularly useful in that regard especially due to its ability to be both an omnibus statistic and provide a glimpse of the variation partition within categories or stages of discrete random variable such as the ones used in this work. Inspecting stage-specific η^2 value, it can be inferred that joint and musculoskeletal degeneration traits (Table 4.2), for instance, are more predictive for Stage 1 with an average value of 0.587 (0.394 to 0.754) compared to Stage 0 - average value 0.373 (0.032 – 0.721). Based on this statistic it can be identified a subset of skeletal traits with highly

predictive Stage 0 traits, the CLV01, RB01, PSY01, PS02, PSY03, S1S2F and IAS01. This subset of consists of skeletal traits for which Stage 0 represents several of the last manifestations of skeletal maturation such as sternal end of the clavicle ossification (CLV01), fusion of the sacral S1 to S2 segment (S1S2F), and late stages of maturation of the symphyseal face of the pubic bone (PSY01 to PSY03). These traits (Stage 0) are key to estimate age for young adults.

Table 4.4 Correlation analysis of skeletal traits with age-at-death for cranial and axial traits.

						Stage η^2		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
CRS01	0.431	0.187	93.618	0.000	500	0.372	0.083	
CRS02L	0.311	0.096	48.124	0.000	500	0.030	0.314	
CRS02R	0.297	0.088	43.990	0.000	500	0.027	0.294	
CRS03	0.411	0.170	84.977	0.000	500	0.162	0.179	
CRS04L	0.497	0.249	124.598	0.000	500	0.341	0.169	
CRS04R	0.518	0.271	135.365	0.000	500	0.373	0.181	
CRS05	0.414	0.173	86.711	0.000	500	0.206	0.140	
CRS06L	0.358	0.129	64.437	0.000	500	0.093	0.184	
CRS06R	0.352	0.125	62.541	0.000	500	0.088	0.184	
C3IS	0.794	0.639	319.268	0.000	500	0.639	0.367	0.838
C4SS	0.810	0.663	331.438	0.000	500	0.704	0.277	0.842
C4IS	0.829	0.695	347.375	0.000	500	0.749	0.275	0.818
C5SS	0.822	0.683	341.531	0.000	500	0.753	0.179	0.793
C5IS	0.831	0.712	355.974	0.000	500	0.829	0.172	0.711
C6SS	0.836	0.714	356.755	0.000	500	0.817	0.108	0.723
C6IS	0.815	0.680	340.129	0.000	500	0.779	0.137	0.677
C7SS	0.798	0.657	328.270	0.000	500	0.760	0.161	0.668
L1IS	0.799	0.646	323.211	0.000	500	0.659	0.238	0.823
L2SS	0.811	0.667	333.621	0.000	500	0.709	0.180	0.794
L2IS	0.835	0.710	355.078	0.000	500	0.782	0.201	0.797
L3SS	0.841	0.719	359.253	0.000	500	0.831	0.062	0.774
L3IS	0.845	0.722	360.765	0.000	500	0.780	0.156	0.813
L4SS	0.844	0.725	362.319	0.000	500	0.862	0.027	0.729
L4IS	0.819	0.678	338.971	0.000	500	0.745	0.105	0.726
L5SS	0.833	0.706	352.905	0.000	500	0.843	0.025	0.688
S1SS	0.801	0.672	335.879	0.000	500	0.825	0.183	0.706
S1S2F	0.554	0.309	154.320	0.000	500	0.943	0.063	

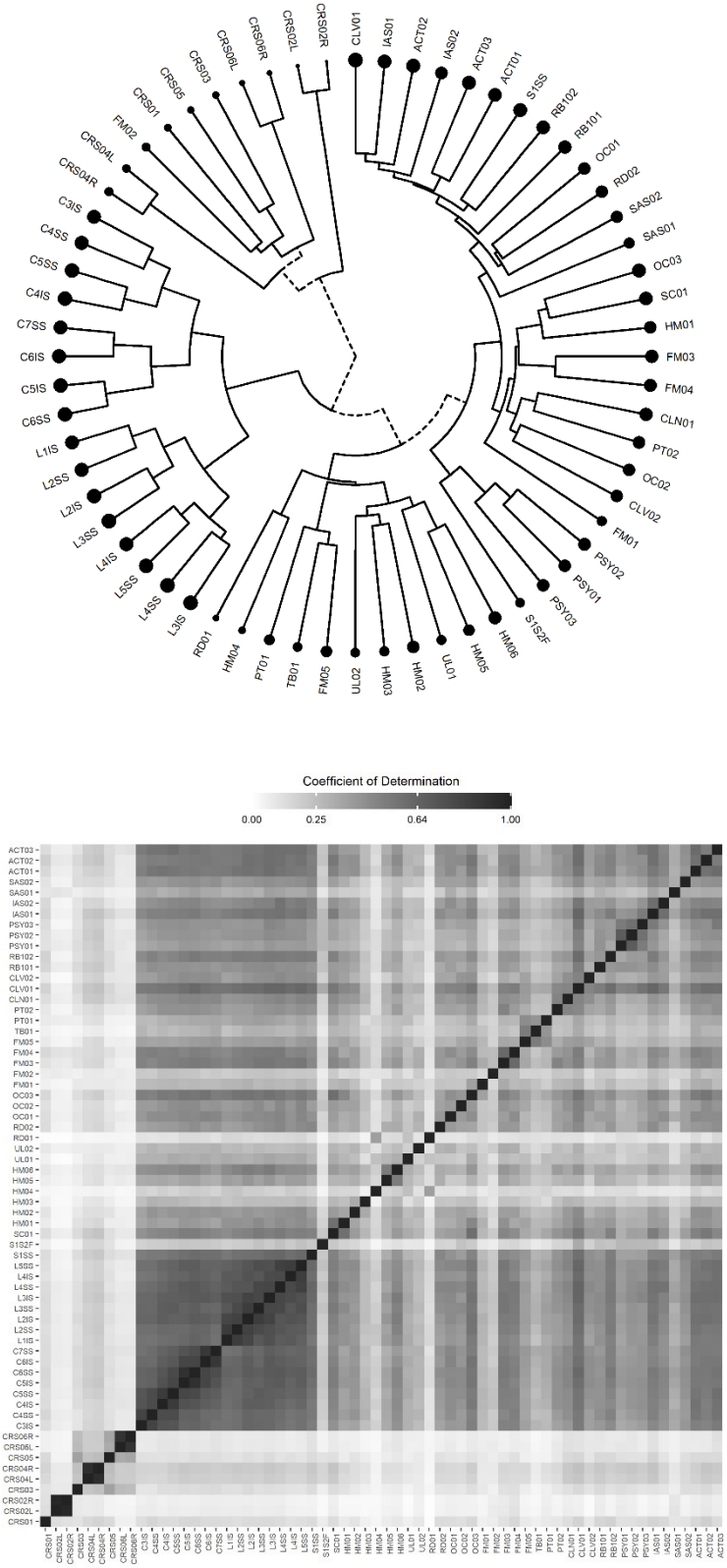

Figure 4.4 Inter-trait coefficient of determination matrix.

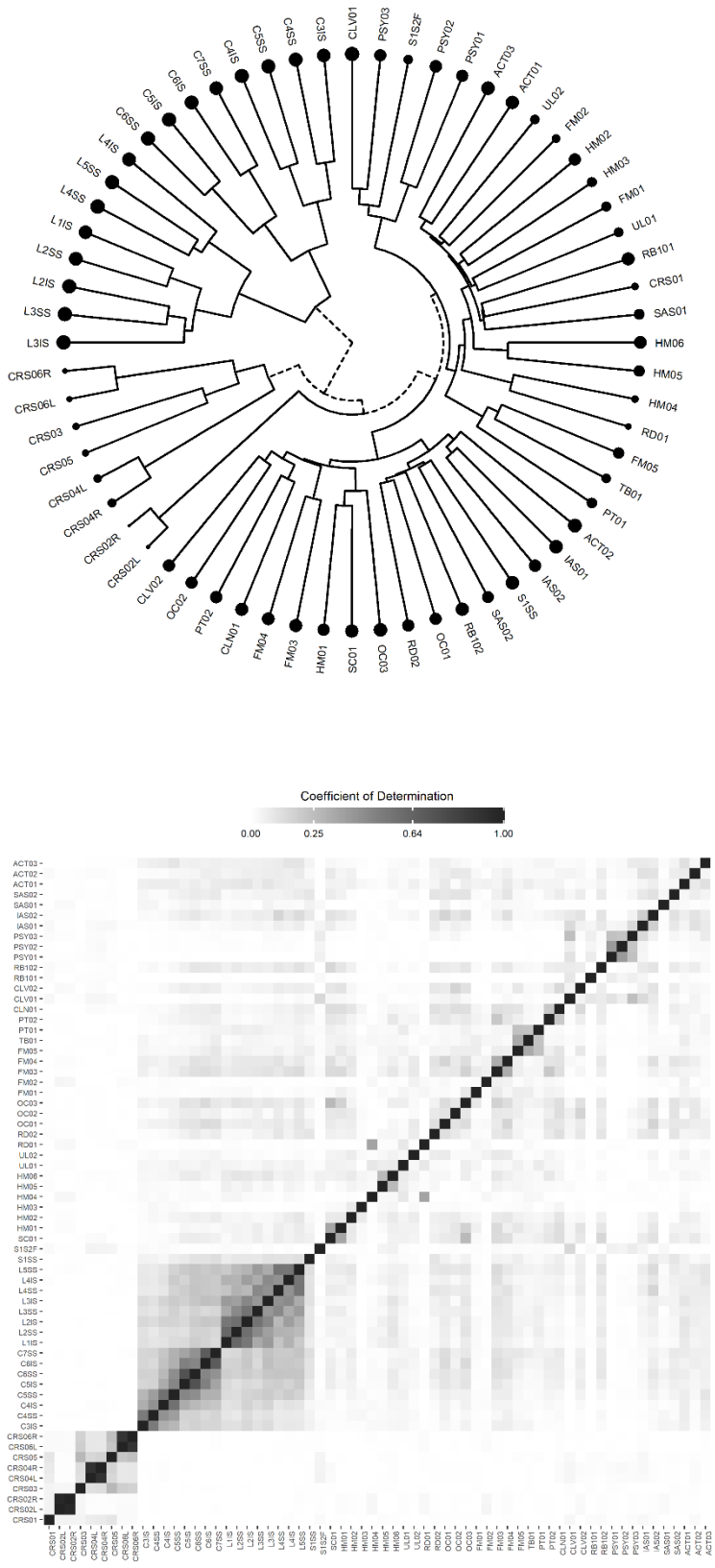

Figure 4.5 Partial inter-trait coefficient of determination matrix controlling for age-at-death.

Figure 4.4 and Figure 4.5 depict inter-trait coefficient of determination matrix, the squared value of Spearman's coefficient, and its natural grouping via hierarchical clustering. The most important point to take from them is the strong inter-trait correlations, even after controlling for age-at-death (partial correlation, Figure 4.5). Partial correlation after controlling for age-at-death, is an important aspect to assess in data processing and age estimation modelling. Strong inter-trait dependencies are useful for data imputation, a maximal signal among traits helps impute latent or missing data with higher accuracy and minimal distortion from the observed data. However, two skeletal traits may be so closely related after controlling for age that knowing the stage or value of one, entirely informs about the value of the other. This is significant because there is no way to improve age inference from a pair of features that are so highly correlated [15]. Accounting for this property of the dataset is key to select an efficient and effective computational modelling approach (Chapter 5). Based on the strong inter-trait dependencies of the dataset, even after controlling for age-at-death, it can be hypothesized that certain modelling techniques are more likely to provide sub-optimal results or ill-conditioned systems such as the case of probabilistic methods under conditional independence assumption or linear regression methods without proper regularization techniques [54, 56, 58, 62, 218].

4.5|Data availability

Data collection in anthropology is time-consuming, often expensive, and bounded by delicate issues concerning ethics and access to human remains or its representation. Data itself represents one of the most valuable assets, particularly in a field where most careers are related to academia and open access to data is the fundamental. Restricted access to data and/or code use for developing or validating models in biological profiling is a prevalent issue as open science and open source initiatives have not been the predominant trait in anthropological research and its outputs [151, 219]. To counteract this tendency, the dataset used in this thesis is available as part of the software developed

as one of its research objectives (Chapter 7). A tab separate value file (.tsv) is on the sub-directory *data-raw* in the code repository of the DRNNAGE software [82]. The data is archived and shared using the open science platform Zenodo [220], can be accessed via *https://www.doi.org/10.5281/zenodo.7274445*. Installing DRNNAGE as a R package enables the dataset to be access in that environment as an object named CAMSAD which stands for Coimbra Adult Macroscopic Skeletal Age Dataset.

5|Age estimation using machine learning: an approach

Machine learning and computational techniques can be used to create predictive models, calculate errors in a statistically sensible way, and compute probability values and other uncertainty quantities that can be presented in court and forensic reports in a rigorous scientific manner [195]. A machine learning based approach is valuable in complex and noisy problems such as skeletal age estimation. Creating explanatory and causal models to map the relationship between skeletal morphology and age-at-death can be an intractable or even an impossible problem to solve. Machine learning algorithms have been used to generate predictive models for age estimation, previous research such as Corsini *et al.* [61], Buk *et al.* [51], Kotěrová *et al.* [52], Stull et al. [221], and Navega *et al.* [32, 81] substantiate how machine learning techniques can be an asset in skeletal age estimation.

Computationally age-at-death estimation can be viewed as function approximation problem. From this perspective, $Y = f(X)$, maps the skeletal traits (X) to an age-at-death (Y). The approach to skeletal age estimation proposed and validated in this thesis relies on a regression-based predictive modelling strategy using machine learning. It consists of using deep random neural networks models to regress age-at-death on skeletal traits coupled with regression uncertainty models to construct predictive intervals.

Current chapter presents the basics on machine learning, and the conceptual and mathematical details of artificial neural networks approach and how predictive intervals can be constructed from a generic framework to address regression uncertainty.

5.1|Machine learning

Machine learning represents a unified algorithmic framework designed to learn and map underlying properties and structural patterns of data, which can be used to describe the data or make predictions and estimates on new data [222–225]. It is subfield of artificial

intelligence, a branch of computer science where areas such computational statistics, applied mathematics, cognitive science, and information theory converge. Machine learning enables performing tasks that would be infeasible to accomplish with human-written, static programs. From a scientific and philosophical standpoint, machine learning is intriguing because understanding it entails a glimpse over the principles underlying human learning and intelligence [226]. It is one of the most prolific research areas in computational sciences due to massive amount of available structured and unstructured data and the relevant problems (*i.e.,* industrial, medical, biological, economic, military) that can be solve efficiently using it [223, 225].

Machine learning tends to differ from traditional statistical modelling. The latter emphasizes modelling accuracy and correctness, that is, creating models that not only reflect the ground truth but are also correct and can provide some explanatory value. Machine learning tends to be more operational and emphasize raw accuracy over understanding the data generation mechanism. Some of the most efficient algorithms in machine learning, i.e., artificial neural networks, random forest, or support vector machines, are usually black boxes from an explanatory perspective contrasting with linear regression models or discriminant analysis. Breiman [227] provides an insightful discussion on these *two cultures* as he referred to it.

Following Jung [225] machine learning can be portrayed as a combination of three components: the data, a model, and a loss function. A plethora of machine learning algorithms results from different choices regarding representation and parameterization of these atomic units. Data is the most important component in a problem to be solve by machine learning[‡].

Data is the collection of data points that contain the information on the features or inputs variables, X, and the label, target or output variable Y. The data itself guides the taxonomy of models and the type of loss function to be a used. Learning from data using machine

[‡] Data as the most important component is one of the reasons why, in this thesis, a chapter was entirely devoted to the dataset constructed and used.

learning generally falls in two main categories referred as unsupervised learning and supervised learning [222]. A third category is reinforcement learning with important applications in robotics and control problems but due to its nature is out of the scope of this overview.

In unsupervised learning do not exist a label or output variable, in such cases the model used aims to discover patterns in the input X. It addresses clustering problems, grouping patterns, and representation problems such as extracting new features from X or reducing dimensionality.

Supervised learning has both X and Y available to the learner or model, which aims to approximate $Y = f(X, \Theta)$. This function maps the relationship between X and Y, Θ represent its learnable parameters. Learning the best parameters is a mathematical optimization problem, that involves minimizing the loss function (Figure 5.1). The loss function, which quantifies the discrepancy between known output and its prediction, varies depending on Y. If the output is numeric the problem can be handled as regression task and the squared error loss is an appropriate function. For classification problems, those with discrete or categorical outputs, the logistic or the hinge loss represent suitable candidates for the loss function [225].

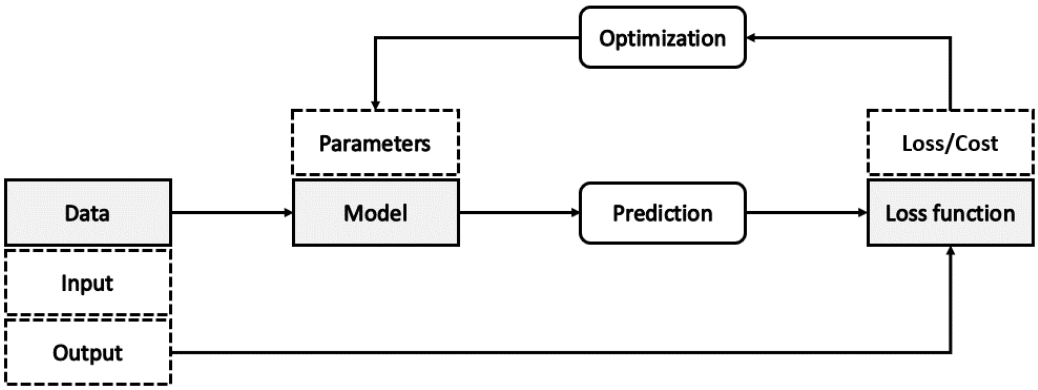

Figure 5.1 Learning as an optimization problem.

5.2|Artificial neural networks

Artificial neural networks (ANN) are a class of connectionist, biologically inspired computational models that enables learning from data for a multitude of tasks such as classification, regression, representation learning, data compression and generation. ANN are function approximation machines and offer a robust and flexible framework for unsupervised, supervised and reinforcement learning. They are in a broad sense the result of two components: architectural bias — that is how many layers and neurons composed the network, and learning strategy — how the parameters of the network are learnt. With the nervous system as its schematic, artificial neural networks are based on interconnected units or nodes also known as artificial neurons. An artificial neuron, Figure 5.2, the basic unit of a network, is a mathematical operator in the form of

$$h(x) = \varphi(\sum_{i=1}^{p} x_i \omega_i + b) \tag{5.1}$$

where $\varphi()$ is an activation function, $\boldsymbol{x_i}$ and $\boldsymbol{\omega_i}$ are the i-th components of the input and weight vector b is the neuron bias. Artificial neurons represent non-linear functions with learnable parameters which ultimately expand this type of model representational capacity to be able to approximate any output function.

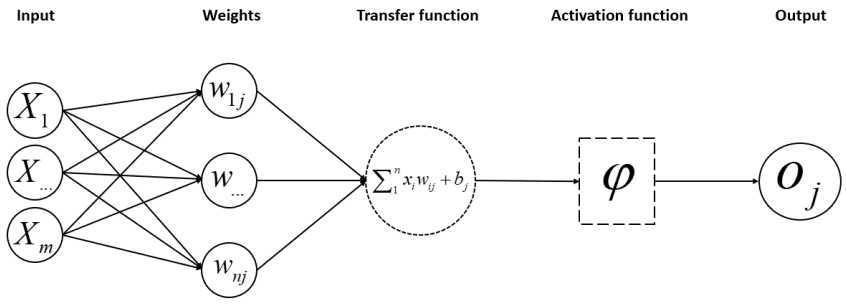

Figure 5.2 Artificial neuron representation.

Figure 5.3 depicts the non-linear behavior of three commonly used activation functions.

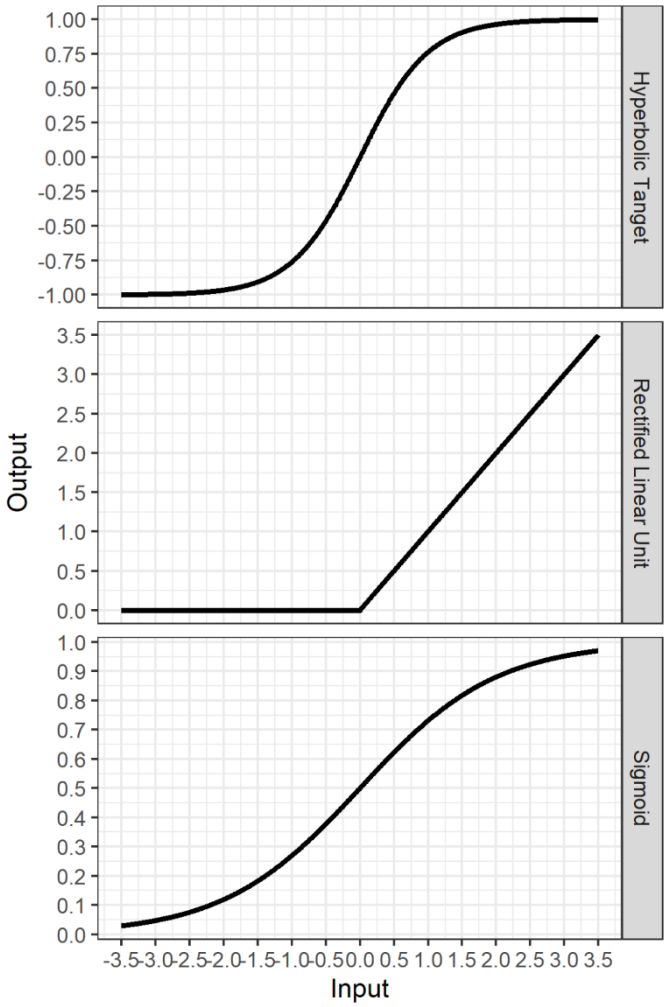

Figure 5.3 Activation function representation.

In its basic implementation, an ANN is composed of three components: the input layer, the 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(s) by means of a non-linear activation function (Figure 5.4.)

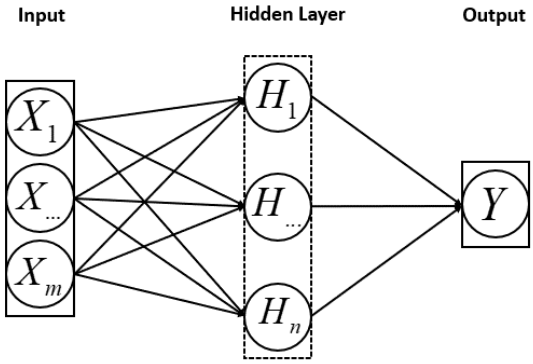

Figure 5.4 Generic representation of a single layer artificial neural network.

A network can represent and map functions of increasing complexity by adding more layers and more nodes within a layer [226]. Networks with multiple layers are usually referred as multi-layer or deep neural networks (Figure 5.5).

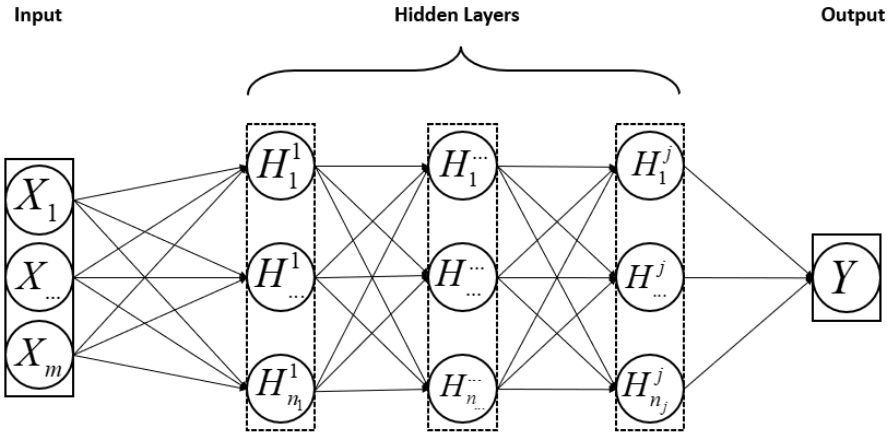

Figure 5.5 Generic representation of a deep (multi-layer) artificial neural network.

A key aspect of ANN is their flexibility and modularity which translate in their capability to be applied to a vast array of heterogeneous data types and domains. The explosion in availability and capacity to store and analyze data in the form of image, video, audio, and unstructured text lead 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 ANN are formulated and trained in the same manner.

A transversal aspect of modern ANN is their use of gradient-based learning algorithms where the weights of a network are iteratively fine-tuned [225, 226]. 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. Genetic programming and evolutionary computing techniques can also be used to train neural network models, particularly to solve the issues related to find the optimal topology of network but also the weights of the network. These approaches are usually referred as neuroevolution, two paradigmatic examples are NeuroEvolution of Augmenting Topologies (NEAT) [228] and Deep Evolutionary Network Structured Representation (DENSER) [229–233].

In this thesis is explored a counterintuitive, yet highly efficient, approach to the training of neural models via randomization in which a subset of parameters - hidden layer(s) weights - are randomly assigned and fixed, and network optimization is recast as least squares estimation problem [234, 235].

5.3|Randomized artificial neural networks

In the context of artificial neural networks, randomization as an intrinsic mechanism of model learning can be traced back to late 80s and early 90s with proposal of randomized *radial basis functions network* (RBF) and the *random vector functional link network* (RVFL) models [236–240]. 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 [241–244]. According to Wang [245], there is no need to rename this strategy for training neural networks, since all key elements have been previously proposed [236–240], 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 (RANN) such as the development of highly efficient algorithms to compute and cross-validate the output layer analytically [246, 247], and its evolution from a framework restricted to shallow networks to a set of techniques and algorithms capable of deep, multi-layered network architectures [248–252].

Fully randomized and fully trainable artificial neural networks represent extremes of a wide algorithmic landscape with a continuum that is only now being explored more thoroughly [235]. While it is difficult to deny the intuitive elegance of fully trainable deep neural networks, the training speed of randomized networks make them a pragmatic choice for many real-world machine learning applications. They can achieve significant accuracy *versus* runtime efficiency tradeoffs from a practical standpoint. This method for training of artificial neural networks may be advantageous for hardware or online implementations [235, 248, 253].

5.3.1|Regularization

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

$$\beta = H^{\dagger} Y \tag{5.2}$$

Where β are the output layer weights, $H^{\dagger}$ is Moore-Penrose pseudo-inverse of the matrix H, which defines the hidden layer, and Y is a column vector storing the network target output. $H^{\dagger}$ can be computed using several methods; a common approach is through orthogonal projection using Eq. (5.3)

$$H^{\dagger} = (H^T H)^{-1} H^T \tag{5.3}$$

From Eq. (5.2) and Eq. (5.3) can be derive that using this algorithm the estimate is obtained as $\hat{Y} = H\beta$ and that the weights of the output layer are in fact the least squares solution that maps the non-linear features induced by the hidden layer of the neural network to its output.

It has been noted 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 [254]. Such 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[255]. Imposing such 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 substituting Eq. (5.3) as follows:

$$H^{\dagger} = (H^T H + \frac{I}{C})^{-1} H^T \tag{5.4}$$

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 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 [246]. Regularization shrinks the

size of the output coefficients towards zeros, which is consistent with the theory that smaller weights result in better generalization of neural networks [255, 256].

An advantage of optimizing the weights the output layer in a RANN solving a least squares estimation problem is the efficient, analytical and closed formulation to assess the leave-one-out (LOO) error as shown by Shao and Er [246] using Allen's [257] Prediction Sum of Squares (PRESS) statistic:

$$E_{LOO} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - hat_{ii}}\right)^2 \tag{5.5}$$

hat_{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. Shao and Er [246] 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 operation is written as $H = U\Sigma V^T$. Using SVD the network estimate can be written as

$$\begin{aligned} \hat{Y} &= H\beta \\ \hat{Y} &= H(H^T H + \frac{I}{C})^{-1} H^T Y \\ \hat{Y} &= U(\Sigma^T \Sigma + \frac{I}{C})^{-1} \Sigma^T U^T Y \end{aligned} \tag{5.6}$$

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 $\left(\Sigma^T\Sigma + \frac{I}{C}\right)\Sigma^T$ 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 Eq. (5.6), that can be obtained from the following Hadamard products (matrix element-wise multiplication)

$$\gamma = U \circ \Gamma^T = U \circ (\Theta \circ U^T) \tag{5.7}$$

with $\Theta = \left(\Sigma^T \Sigma + \frac{I}{C} \right)^{-1} \Sigma^T$. The diagonal elements of the projection matrix, hat_{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

$$\hat{y}_i = \frac{y_i - f(x_i)}{1 - hat_{ii}} \tag{5.8}$$

In addition to this highly efficient computational formulation to train and regularize a randomized neural network, data standardization and the addition of gaussian noise to the several components of the network can also improves generalization.

5.3.2|Deep random neural networks

The mathematical and network formulation presented so far pertain to a randomized single hidden layer network architecture. However, several authors proposed different techniques to extend RANN to deeper architectures [248–252]. To increase the depth of the network one can, resort to fully randomized approaches or use autoencoding strategies and stack multiple autoencoding RANN to build a multi-layer or deep network. In this thesis both approaches are followed.

The fully randomized approach pursued follows the proposal of Shi *et al.* [252] to train deep randomized neural network models (DRNN) due to its simplicity. Following the

authors, the first layer of the network is defined as

$$H^{(1)} = g(XW^{(1)}) \tag{5.9}$$

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

$$H^{(j)} = g(H^{(j-1)}W^{(j)}) \tag{5.10}$$

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

$$H^{(j)} = g([H^{(j-1)}\ X]W^{(j)}) \tag{5.11}$$

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

$$D = \left[H^{(1)}\ H^{(2)} \ldots H^{(j-1)}\ H^{(j)}\ X \right] \tag{5.12}$$

The design of the deep network using this formulation is very similar to a shallow RANN and 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 links or skips layers. That is, as aforementioned, the key difference between and ELM and RFVL networks.

The autoencoding strategy adopted is based on the supervised autoencoding approach proposed by Tissera and McDonnell [248]. The authors propose an algorithm for

deep randomized neural network for image classification, but it is generalizable to regression problems and tabular data as performed in this thesis. This deep random neural network is trained with same mechanism already presented for efficient regularization and analytical cross-validation, nonetheless it introduces some conceptual and notation changes that will be briefly addressed.

In this approach, the depth of the network is increased by stacking autoencoders. An autoencoder is a neural network that attempts to reproduce its input as its output. It has an internal layer that describes the code used to represent input. Typically, they are constrained in such a way that they can only reproduce inputs that closely resemble the training data. Because the model is required to prioritize which aspects of the input should be replicated, it frequently discovers useful data properties. Autoencoders discover encoding and decoding functions. Autoencoders with nonlinear encoder functions and nonlinear decoder functions are therefore capable of learning a more potent nonlinear generalization of principal component analysis (PCA) [226]. A randomized autoencoder learns to recover the input from its distorted version as encoded by the hidden layer with random weights.

Following Tissera and McDonnell [248] rather than directly predict Y from X, a randomized autoencoder is trained to recover and reconstruct $S=[X\,Y]$, the concatenated version of the input and output, from the randomized hidden layer that act as an encoder. Because the output Y is in fact quantity to be predicted, the input to the first layer of a deep random neural network trained with this approach is substituted by $S_{j=1}=[X\,Y_0]$, where Y_0 is a vector of zeros or other meaningful value (i.e. average of Y) that acts as placeholder of Y in the decoder layer. The input to subsequent layers is $\bar{S}_j=[\bar{X}_{j-1}\,\bar{Y}_{j-1}]$, the autoencoded version of S. The autoencoded data is the input of another random autoencoder with S as its output. The striking feature here is that $\bar{X}_j$ is the recovered or reconstructed version of X but $\bar{Y}_j$ represents a prediction or estimation of Y because it was not available to the encoder layers due to its substitution by Y_0 in the first layer.

Throughout the depth of the network there is an auto-corrective behavior that

refines the prediction of Y while minimizing the reconstruction error over X at the same time. By stacking autoencoders in this fashion, the network aims to improve its autoencoding of S, and implicitly performs a prediction of Y as $\hat{Y}=\overline{Y}_j$ at every layer along the depth of the network hence the supervised aspect of this autoencoder which are typically trained in an unsupervised fashion – only in respect to X.

5.3.3|Implicit ensemble models

One key advantage of the randomized approach used in this thesis is that it can enable implicit neural ensemble models [252]. For the fully randomized approach it means that rather than applying Eq. (5.2) once to solve the output layer weights (solution), Eq. (5.2) can be re-used along the depth of the network for each $H^{(j)}$ computed from Eq. (5.9) or Eq. (5.10) and obtain estimates. The final estimate can be then obtained by averaging all estimates along the network depth. In the case of the supervised autoencoding strategy, $\overline{S}_j$ is averaged along the depth of the network. This feature stabilizes the predictions and offer a different mechanism to train an ensemble model other than training each model independently.

5.4|Regression uncertainty modelling

Previous section presented the foundation for mathematical estimation using neural networks as a numeric prediction or regression problem. Yet, it focused only on how point estimates can be obtained, that is the conditional expectation of Y given the observed values X. Mapping the uncertainty of the point estimate is essential in several domains, which means that a predictive interval for a prescribed confidence level should also be part of the analysis and the subsequent report. Incorporating prediction uncertainty into deterministic estimates improves the model's credibility and reliability [258].

In this thesis a generic approach based on recast uncertainty estimation as

regression problem was pursued. Simply put this approach consists of regressing absolute residuals, (estimation errors), on the values of the of estimate itself. This is formally defined as

$$\hat{E}(\hat{Y}) = \beta_0 + \beta_1 \hat{Y}^{\eta} \tag{5.13}$$

where $\hat{E}(\hat{Y})$ is the estimate absolute residual based of existing estimate, $\hat{Y}$ [§], β_0 and β_1 are the model coefficients and η is a exponent parameter that introduces non-linearity in the model. If $\eta = 1$, the model operates in a linear fashion. Estimation error or absolute residuals are defined as $E = \left| Y - \hat{Y} \right|$. The model formulated by Eq. (5.13) is from herein referred as a regression uncertainty model (RUM). This approach is based on Milborrow [259] and was selected due to its conceptual simplicity and flexibility as it can be used to infer prediction uncertainty for any underlying regression algorithm that generates $\hat{Y}$. The estimate of $\hat{E}(\hat{Y})$ is fundamental for two strategies, parametric and non-parametric, used in this thesis to infer predictive intervals. A parametric solution is based on assuming that uncertainty as a distribution with a gaussian shape around the point estimate. The non-parametric, distribution-free is based on conformal prediction framework [260–262].

5.4.1|Truncated gaussian

A regression uncertainty model can be used to output a gaussian shaped predictive interval. In this strategy the predicted residual, $\hat{E}(\hat{Y})$, is first scaled by 1.2533. The scaling factor is the ratio of the standard deviation to the absolute deviation [125, 126]. Assuming normality on the variance around each point estimate, the prediction interval associated to a regression model is given by the quantiles of a gaussian or truncated gaussian

[§] In this section its assumed that the estimates, $\hat{Y}$, used in the RUM are obtained using some form of cross-validation.

parameterized with the conditional mean $\hat{Y}$, and conditional variance $\hat{Y}_{\sigma} = 1.2533\hat{E}(\hat{Y})$. A truncated gaussian distribution is favored as it guarantees that the prediction interval is within the domain of the problem at hand. The key advantage of this approach is its simplicity compared to likelihood methods [15–17, 20, 23, 127–129]. In addition to the numerical interval, this approach also allows visualization as illustrated by Figure 5.6.

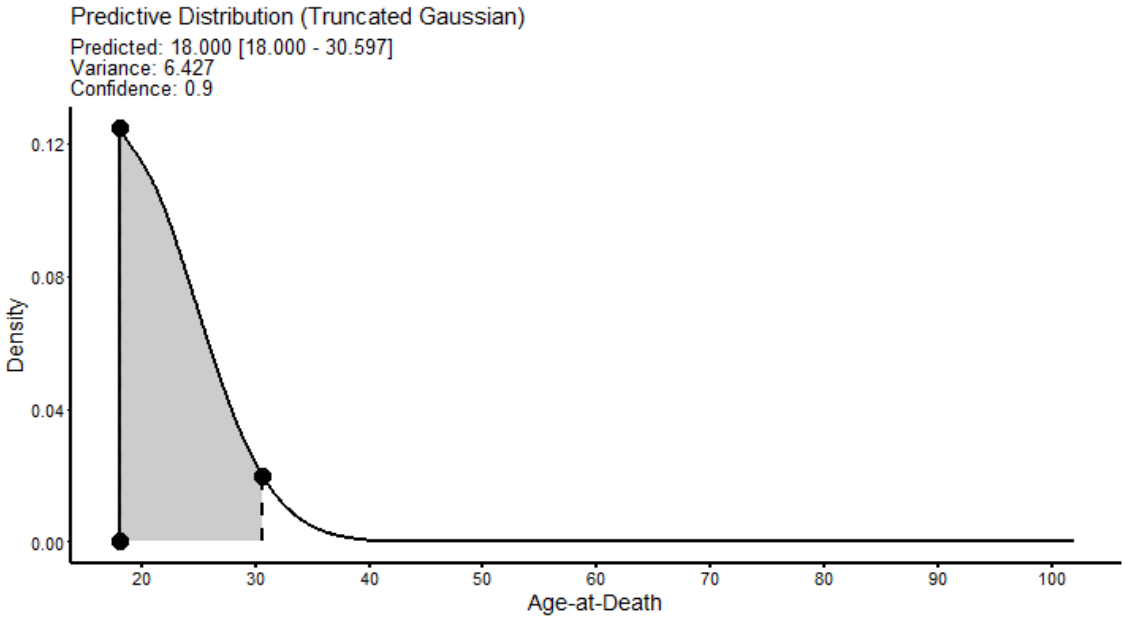

Figure 5.6 Prediction interval using a gaussian uncertainty model in age estimation problem.

5.4.2|Conformal prediction

Conformal prediction (CP) is a learning paradigm and framework proposed by Vovk et al. [260] to complement the estimates and predictions of machine learning algorithm with confidence measures. The following key points characterize this approach: it gives provably valid measures of confidence; the only assumption it makes about the data is that they are independently and identically distributed (i.i.d); it can be adapted to accommodate any learning algorithm; and it can be used in classification and regression settings. Conformal prediction uses past experience, i.e. error patterns of a machine

learning algorithm to build distribution-free confidence measures [247, 262–267].

Introducing a training set as $\mathbf{z}^l = \{\{\mathrm{x}_i, \mathrm{y}_i\},\ i = 1,2,\ldots,\mathrm{l}\}$ where x_i denotes the input vector, and y_i denotes the output, both instances of X and Y respectively. Given a testing example x_{l+1}, the objective of conformal prediction is to use the information of the training set to construct a predictive region that contains the known output, y_{l+1}, with a high and preset level of probability, i.e. 0.9 or 0.95. Assuming only that the data is drawn from the same (unknown) distribution, a conformal predictor tests every possible $\hat{y}$, predicted value, for each augmented dataset $z^l \cup \{x, \hat{y}\}$, and $l + 1$ nonconformity scores are computed as:

$$\begin{aligned} \alpha_i^{\hat{y}} &= A(\{(x_1, y_1),\ldots,(x_{(i-1)}, y_{(i-1)}),(x_{(i+1)}, y_{(i+1)}),\ldots,(x_{(l+1)}, \hat{y}_{(l+1)})\},(x_i, y_i)) \\ \alpha_{(i+1)}^{\hat{y}} &= A(\{(x_1, y_1),\ldots,(x_l, y_l)\},(x_{(l+1)}, \hat{y}_{(l+1)})) \end{aligned} \tag{5.14}$$

with $i = 1, 2, \ldots, l$. $A(S, z)$, a nonconformity measure, defines a function measuring the degree of disagreement between observation z and the dataset S. In regression problems the nonconformity function is typically defined as the absolute error

$$A(S,z) = |\, y - \hat{\mu}(x) \,| \tag{5.15}$$

where $\hat{\mu}$ is the regression function built from, S, the training set and $\hat{\mu}(x)$ is the estimate generated by it. By using *A(S, z)* it is possible to compute a p-value for every $\hat{y}$ which indicates for a given datum the probability of $\hat{y}$ being the correct predicted value. Such p-value is calculated as:

$$p(\hat{y}) = \frac{|\{i \in \{1,2,\ldots,l\} \,|\, \alpha_i^{\hat{y}} \geq \alpha_{(l+1)}^{\hat{y}}\}| + 1}{l+1} \tag{5.16}$$

with the p-values calculated with Eq. (5.16) and a preset *significance level*, α, a predictive region that contains the known output with $1-\alpha$ confidence is $\{\hat{y} : p(\hat{y}) > \alpha\}$. We refer here to predictive region as a more generic term that includes the notions of predictive set

and predictive interval. CP is also used to provide confidence measures to classification algorithms and defining predictive set of possible predicted classes for a *datum*.

The notation of CP might seem daunting at first because it assumes a transductive or *on-line* approach to learning. That is, predictions are formed in sequence, one at a time using previous experience to define predictive regions and confidence measures. This property gives solid theoretical and empirical validity to the prediction region obtained but it might be computationally prohibitive to implement in practice. CP is built on top of a machine learning algorithm usually referred to as the *underlying algorithm*. Depending on the complexity of the algorithm or the volume of data it might not be cost effective to apply transductive conformal prediction.

To tackle this problem, techniques that explore and exploit data splitting and cross-validation strategies have been proposed [262, 265, 266, 268]. Cross-Conformal Prediction (CCP) and Jackknife (or Leave-One-Out) Conformal Prediction (JCP) are two variants used to implement CP in a computationally efficient manner. Both techniques use cross-validation to compute the nonconformity measure, *A(S, z),* needed in CP. One important remark is that CCP and JCP are only empirically valid [262, 266]. That means that their finite out-of-sample coverage probability in regression problems is not theoretically guaranteed to be $1-\alpha$ but empirical studies on hypothetical and real problems demonstrated that the coverage probability converges to the preset level of confidence.

In practice, in a regression problem it is not feasible to apply Eq. (5.14) and Eq. (5.16) to every possible predicted value to compute the conformal p-value and to construct the prediction interval. Inferring the nonconformity function is instead recast as regression problem and is applied a regression algorithm to learn the nonconformity function based on the absolute residuals obtained from cross-validation on the training set. This is already accomplished via the regression uncertainty model defined in Eq. (5.13).

Computing a predictive interval using conformal prediction involves a scaling operation analogous to the one involved in the truncated gaussian strategy. However, this scaling factor is given as

$$CF_{\alpha} = Q_{1-\alpha}\left(\frac{\left|Y-\hat{Y}\right|}{\hat{E}(\hat{Y})}\right) \tag{5.17}$$

where α confidence level, $Q_{1-\alpha}(.)$ is a quantile function that returns the $1-\alpha$ quantile of the ratio between the observed absolute residuals and the estimated absolute residuals as computed by the RUM, Eq. (5.13). The predictive interval is then obtained as $\hat{Y} \pm CF_{\alpha}\hat{E}(\hat{Y})$, truncation of the lower and upper values obtained can be performed for assurance that the predictive interval is consistent with range of Y for a given problem.

6|Analysis of machine learning models for age estimation

Current chapter constitutes the experimental core of this thesis and the provides the results of two *in silico* experiments, that is by means of computational analysis and simulation. These experiments were conducted to assess and validate the machine learning approach described in Chapter 5 for constructing age estimation models using deep random neural networks and regression uncertainty modelling and test the hypothesis that multifactorial age estimation models provide more accurate age estimates. This chapter details the computational procedure and metrics used to assess model performance, describes the model parameterization, and discusses the results obtained.

6.1|Cross-validation scheme

To assess the performance of DRNN and RUM models in multifactorial age estimation from macroscopic skeletal traits was followed a simple template for robust metric assessment based on a resampling method known as Monte Carlo cross-validation (MCCV) or repeated random sub-sampling validation [269]. It 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 Eq. (5.5 to 5.7) to optimize the regularization parameter C and obtain leave-one-out predictions, which are used to train the RUM models. C is optimized as 2^x with $x \in \{-6,-4,...,12\}$. 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 train 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 the computational analysis is organized in two experiments, from now on referred as experiment A and B:

A) The first experiment 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, models were fitted to different anatomical complexes or set of traits that mimic existing aging standards—i.e., a model for the sutures or the pubis symphysis.

B) 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 95% predictive intervals (95% PI) were computed by setting the uncertainty to parameter $\sigma = 0.05$. DRNN models were constructed using the fully randomized and the supervised autoencoding approaches described in Chapter 5, both models are built as implicit ensembles. Fully randomized approach used direct links connecting the input to every subsequent layer. RUM models were built using both approaches previously described and operated in a linear fashion $\eta = 1$, Eq. (5.13). More details on model parameterization are provided in section 6.3.

6.2|Metrics

In this section are described the mathematical descriptors used to assess model performance in age estimation. An age-at-death estimation model — regardless of its underlying mathematical algorithm — should be accurate, unbiased, valid, and efficient. Accuracy refers to ability 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

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{6.1}$$

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 omnibus way to assess bias ($\hat{\beta}_e$) is by computing the slope of regression line of the residuals, $e_i = y_i - \hat{y}_i$, on known values. When minimal to no bias is presented, this value should be close to zero. A positive slope suggests a systematic bias as the one describe previously. Bias is computed in as

$$\hat{\beta}_e = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(e_i - \bar{e})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6.2}$$

where $\bar{y}$ and $\bar{e}$ are the means of the known and residuals values.

The validity of model, in the context of this study refers to the ability to contain the known age within the predictive interval within a reasonable margin close to the nominal uncertainty level allowed. For an uncertainty level (α) of 0.05 (or 5%) is expected that the coverage, correct proportion of individuals within the predictive interval, is close to 0.95 (or 95%). This validity measure is computed as

$$P(\alpha) = \frac{\sum_{i=1}^{n} \delta(y_i, l_i, u_i)}{n} \tag{6.3}$$

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 $\delta(y_i, l_i, u_i) = 0$ otherwise. l_i and u_i are the values of the lower and upper ends of the

predictive interval respectively.

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

$$PIW = Q(u - l, \tau) \tag{6.4}$$

where $Q(.)$ is a quantile function and τ a given quantile, with $\tau \in \{0.5, 0.025, 0.975\}$. This computes the median of the predictive interval width and its associated 95% confidence interval (quantile-base).

6.3|Parameterization

A key aspect of any ANN model is its architecture, that is how many neurons (or nodes), and layers compose the network. To leverage the full potential of DRNNs, maximize its training speed and efficiency rather than search for the optimal architecture, a simple heuristic was developed based on 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, is proposed the following heuristics for setting the architecture of a DRNN as built in this thesis to handle tabular data. The width, size, or number of neurons of each layer was set as

$$S = 2^{\left\lfloor \log_2\left(8\sqrt{\frac{2^k}{k}}\right)\right\rfloor}, k = \log_2(n) \tag{6.5}$$

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

$$L = 2^{\lfloor \log_2(k) \rfloor}, k = \log_2(n) \tag{6.6}$$

Following Eq. (6.1) and Eq. (6.2) as a simple heuristic allows predictable, parsimonious network architectures. This way the network permits many computing units for randomized feature extraction distributed along several layers without incurring in overparameterization. This heuristic also leverages the simplicity of training a deep (randomized) neural network using the same mechanisms of a shallow one, while exploiting implicit ensemble framework. Applying the described heuristic defines the network architecture with a *rectangular* topology composed by eight layers of 32 neurons each, for a total of 256 randomized units.

DRNN are computationally cheap nonlinear models built by combining regularized linear regression with nonlinear features obtained by using an activation function, $\varphi()$, with random weights. In this work, the rectified linear unit (ReLU) is used as the nonlinearity of the networks. The ReLU is defined as $\varphi(z, w) = \max(0, zw)$, where z and w are the layer input and random weights matrices. Since the regularization process involved in the training process described in this work is not scale invariant, Eq. (5.6), during network training normalization by mean centering and variance scaling was performed on the matrices X, XW, H and Y. The output of the network was later rescaled before computation of the performance metrics. Scaling the output translates in supervised autoencoding approach as substituting Y by Y_0 in the first autoencoder as the mean known average of Y.

Network 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 Chapter 7). 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.

6.4|Results

Results from experiment A are condensed in Table 6.1 to Table 6.4 for both types of deep random neural network and regression uncertainty models. From a technical and computational perspective, fully randomized and supervised autoencoding architectures provide very similar age estimates. However, were detected some key differences as reflected in the metrics presented. Mean absolute error (MAE) presented a relative absolute difference of approximately 2% between models using a fully randomized approach and autoencoding strategy. A similar pattern was observed for validity metric, $P(\alpha)$, and efficiency as measured by the median predictive intervals width (PIW) for both uncertainty estimation models — using a gaussian or conformal prediction approach. Neural network models presented a relative absolute difference of approximately 52% on the estimation bias metric, $\hat{\beta}_e$. Supervised autoencoding models (Table 6.3 and Table 6.4) presented a two-fold reduction in bias compared fully randomized models, which despite by a marginal amount produced more accurate models (Table 6.1 and Table 6.2). Regression uncertainty models differ marginally, 1.6%, regarding the approach used. Both approaches produced valid models converging to expected nominal coverage, 0.95, with a global average of 0.947.

From an anthropological perspective it was observed that models based solely on the cranial sutures exhibited the worst performance among all models produced with a MAE of 15.754 (13.586-18.386) and a median predictive interval width (PIW) of 67.032 years, which renders the cranial sutures an inaccurate and inefficient set of traits. Modelling based on other specific anatomical regions resulted in DRNN models with an average median MAE of 9.062 (7.583-11.560), focusing solely on this metric it is reasonable to state that on its own different anatomical regions perform similarly in age estimation based on the results presented. 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, reducing both mean absolute error and bias (Figure 6.1). 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 (5.561–7.781, 95% CI) and decreased the prediction bias considerably when compared to any model built on the same anatomical regions independently, and a PIW of 33.650 (11.267–41.087, PIW 95% CI). A model based only on degenerative traits (m = 39) resulted in a MAE of 6.938 (6.010–7.896, 95% CI) and median PIW of 33.012 (26.827–35.122, PIW 95% CI).

From the results presented multifactorial age estimation models clearly provide improved efficiency as reflected in narrower predictive intervals. From Figure 6.2 to Figure 6.4 it can be also observed that multifactorial models provide accurate and efficient estimates across the entire adult lifespan, solving the problem of open-end and unspecific age-at-death estimates for the elderly.

Figure 6.2 illustrates the importance of non-standard traits to predict accurately advanced age-at-death. Based solely on degenerative traits of the vertebrae, limb joint and musculoskeletal attachments sites we can obtain estimates for the elderly that are comparable to more classical traits (Figure 6.3) or full set models (Figure 6.4). The downside of relying solely on this type on indicators 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 5.877 (4.970–6.728, 95% CI), and PIW of 29.231 (12.619–37.876, PIW 95% CI) years. The prediction bias for this model was 0.091 (0.002–0.170, 95% CI), which represents a two-to-six-fold reduction in the prediction bias compared to other models built on specific anatomical regions individually.

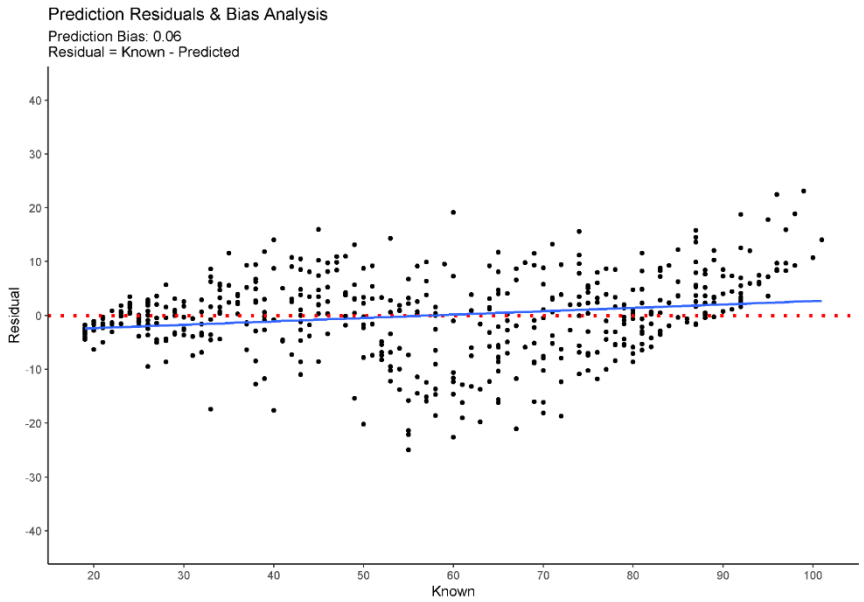

Figure 6.1 Bias analysis of full multi-trait model. Deep supervised autoencoder model.

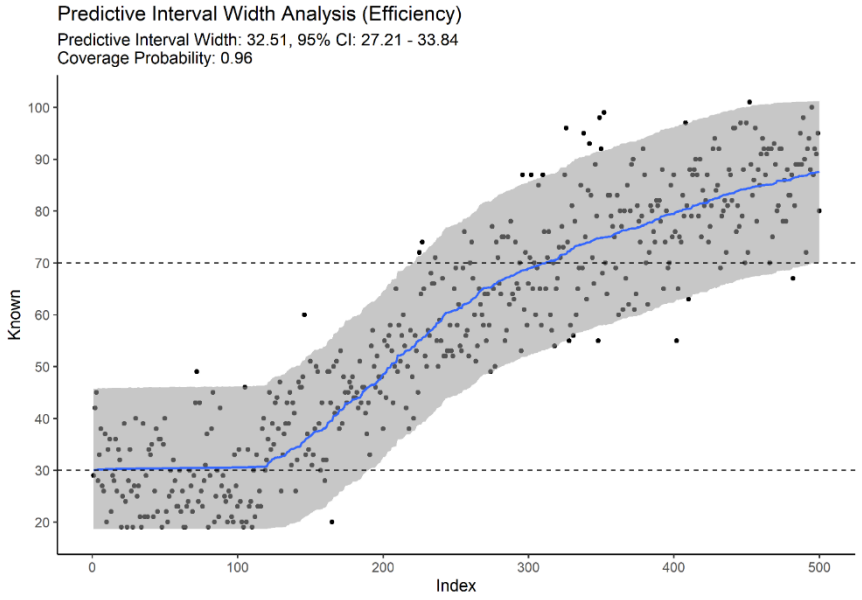

Figure 6.2 Predictive efficiency of degenerative traits of the axial and appendicular skeleton, $\alpha = 0.05$. Fully randomized model with pooled uncertainty models.

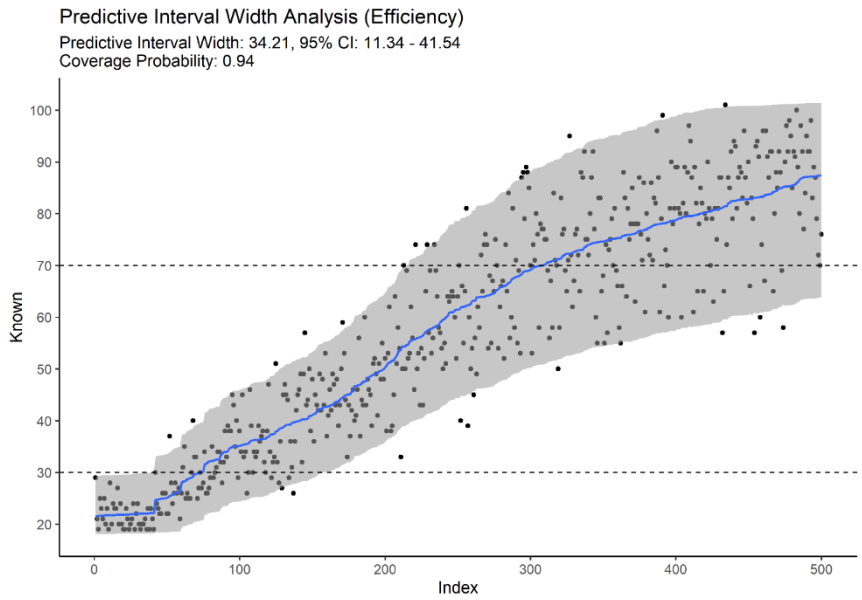

Figure 6.3 Predictive efficiency of standard age-related traits (Clavicle, 1st rib, pubic symphysis, sacroiliac joint, S1S2 fusion and acetabulum), $\alpha = 0.05$. Fully randomized model with pooled uncertainty models.

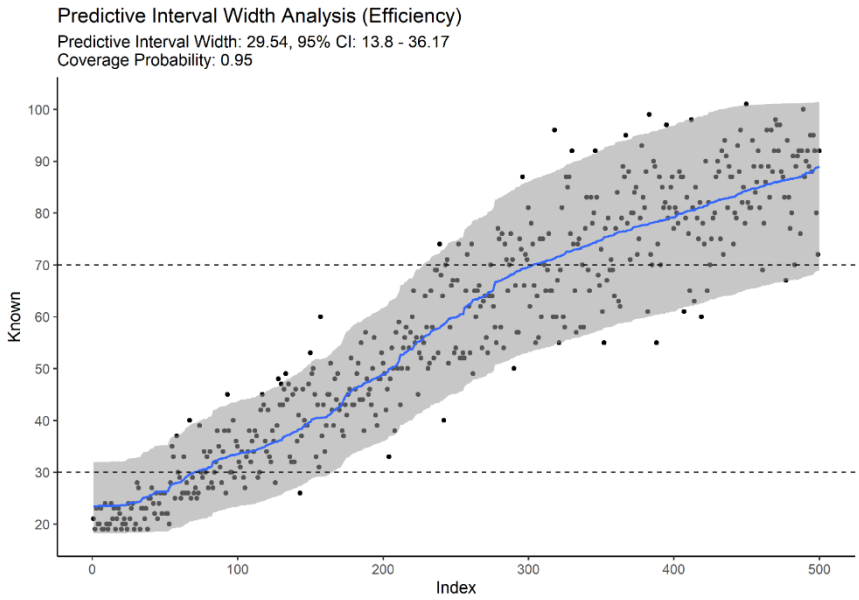

Figure 6.4 Predictive efficiency of full multi-trait model, $\alpha = 0.05$. Fully randomized model with pooled uncertainty models.

Table 6.1 Monte Carlo cross-validation for models built on pre-specified skeletal traits sets. Ensembled deep randomized neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.300	0.656	0.950	68.144	51.699	69.759
(m = 9)	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	Median	8.185	0.198	0.960	38.754	33.732	40.842
(m = 16)	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	Median	7.583	0.167	0.960	37.378	29.109	39.541
(m = 16)	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	Median	8.949	0.244	0.960	49.234	17.354	51.610
(m = 2)	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	Median	9.500	0.277	0.950	48.936	24.334	49.637
(m = 2)	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	Median	10.897	0.370	0.940	51.210	26.905	56.954
(m = 3)	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	Median	8.523	0.223	0.950	44.668	20.378	47.969
(m = 4)	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	Median	8.886	0.229	0.970	42.978	31.727	45.742
(m = 3)	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	Median	6.962	0.147	0.970	33.732	28.882	35.122
(m = 39)	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	Median	6.609	0.147	0.950	34.245	12.927	41.087
(m = 16)	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
Full	Median	5.925	0.117	0.950	30.010	15.631	36.081
(m = 64)	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 6.2 Monte Carlo cross-validation for models built on pre-specified skeletal traits sets. Ensembled deep randomized neural network with conformal prediction regression uncertainty model.

				Conformal prediction			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.300	0.656	0.950	68.900	49.040	71.236
(m = 9)	95% CI	13.586	0.590	0.890	66.060	43.889	68.577
		17.206	0.732	0.990	70.998	53.274	73.619
Axial	Median	8.185	0.198	0.950	37.194	34.066	38.449
(m = 16)	95% CI	7.365	0.137	0.900	35.626	32.516	36.681
		9.139	0.260	0.990	39.218	35.256	41.154
Appendicular	Median	7.583	0.167	0.950	35.784	29.844	39.016
(m = 16)	95% CI	6.678	0.103	0.890	33.497	28.391	36.832
		8.523	0.231	0.990	37.873	31.269	40.998
Clavicle	Median	8.949	0.244	0.950	45.026	16.586	50.522
(m = 2)	95% CI	7.798	0.169	0.890	34.735	15.362	47.707
		10.192	0.307	0.990	52.211	17.929	52.799
First rib	Median	9.500	0.277	0.950	48.682	24.403	52.042
(m = 2)	95% CI	8.138	0.204	0.890	45.774	22.379	49.911
		10.831	0.351	0.990	50.630	26.340	53.750
Pubic symphysis	Median	10.897	0.370	0.950	54.089	27.605	62.231
(m = 3)	95% CI	9.371	0.280	0.890	51.271	25.070	59.326
		12.542	0.459	0.990	58.940	29.789	64.487
Sacroiliac complex	Median	8.523	0.223	0.950	44.205	20.725	49.184
(m = 4)	95% CI	7.380	0.145	0.890	38.758	18.815	47.366
		9.742	0.288	0.990	47.599	22.472	52.106
Acetabulum	Median	8.886	0.229	0.950	40.346	31.292	44.148
(m = 3)	95% CI	7.758	0.162	0.890	37.814	29.581	41.740
		10.006	0.287	0.990	43.004	32.866	46.194
Degenerative traits	Median	6.962	0.147	0.950	32.091	29.553	33.650
(m = 39)	95% CI	6.084	0.085	0.890	30.592	28.172	31.642
		7.814	0.200	0.990	33.607	30.783	35.527
Standard	Median	6.609	0.147	0.950	33.992	13.391	43.516
(m = 16)	95% CI	5.561	0.087	0.890	29.790	12.323	41.647
		7.598	0.202	0.990	37.381	14.785	45.398
Full	Median	5.925	0.117	0.950	29.441	16.425	37.510
(m = 64)	95% CI	5.101	0.060	0.900	25.871	15.199	35.523
		6.728	0.170	0.990	32.573	17.685	39.171

Table 6.3 Monte Carlo cross-validation for models built on pre-specified skeletal traits sets. Deep supervised autoencoder neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.981	0.449	0.920	63.107	42.025	69.148
(m = 9)	95% CI	13.796	0.311	0.860	60.160	38.291	67.163
		18.360	0.598	0.970	65.501	45.677	70.694
Axial	Median	8.207	0.111	0.940	36.809	31.336	39.567
(m = 16)	95% CI	7.179	0.033	0.880	34.051	29.891	37.953
		9.326	0.177	0.980	38.675	32.600	41.163
Appendicular	Median	7.736	0.100	0.950	36.191	27.618	38.889
(m = 16)	95% CI	6.825	0.025	0.900	34.533	26.255	36.919
		8.796	0.167	0.990	38.052	28.886	40.703
Clavicle	Median	9.231	0.149	0.950	48.670	16.238	49.510
(m = 2)	95% CI	8.011	0.065	0.910	39.946	15.008	48.134
		10.511	0.235	0.990	50.708	17.482	50.995
First rib	Median	9.635	0.165	0.950	46.990	21.762	49.399
(m = 2)	95% CI	8.412	0.075	0.900	45.252	20.308	47.552
		11.121	0.246	0.990	48.500	23.203	51.141
Pubic symphysis	Median	11.560	0.227	0.940	53.625	27.090	56.460
(m = 3)	95% CI	10.111	0.109	0.880	50.240	24.910	54.635
		13.193	0.349	0.980	55.817	29.210	58.184
Sacroiliac complex	Median	8.575	0.132	0.930	43.663	17.616	46.552
(m = 4)	95% CI	7.379	0.052	0.860	38.757	16.069	44.800
		9.875	0.208	0.980	45.756	19.101	48.290
Acetabulum	Median	8.845	0.134	0.950	41.166	28.798	45.032
(m = 3)	95% CI	7.660	0.056	0.900	39.360	27.090	43.341
		10.053	0.199	0.990	42.778	30.236	46.576
Degenerative traits	Median	6.915	0.080	0.940	32.293	26.827	33.890
(m = 39)	95% CI	6.010	0.004	0.880	30.923	25.468	32.117
		7.896	0.143	0.980	33.583	28.094	35.604
Standard	Median	6.650	0.085	0.930	33.054	11.267	39.419
(m = 16)	95% CI	5.739	0.021	0.870	29.418	10.224	37.585
		7.781	0.147	0.980	36.742	12.593	41.154
Full	Median	5.829	0.064	0.930	28.452	12.619	34.482
(m = 64)	95% CI	4.970	0.002	0.860	25.387	11.187	32.688
		6.686	0.119	0.970	31.534	14.876	36.273

Table 6.4 Monte Carlo cross-validation for models built on pre-specified skeletal traits sets. Deep supervised autoencoder neural network with conformal prediction regression uncertainty model.

				Conformal prediction			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.981	0.449	0.940	67.976	40.592	79.931
(m = 9)	95% CI	13.796	0.311	0.870	63.536	36.442	75.268
		18.360	0.598	0.980	71.779	44.579	84.595
Axial	Median	8.207	0.111	0.940	38.285	32.327	40.240
(m = 16)	95% CI	7.179	0.033	0.880	35.451	30.721	37.626
		9.326	0.177	0.980	40.496	33.760	42.941
Appendicular	Median	7.736	0.100	0.940	36.135	28.168	39.193
(m = 16)	95% CI	6.825	0.025	0.880	33.701	26.688	36.449
		8.796	0.167	0.990	38.620	29.544	41.816
Clavicle	Median	9.231	0.149	0.950	46.688	14.427	51.192
(m = 2)	95% CI	8.011	0.065	0.890	37.260	13.191	48.577
		10.511	0.235	0.990	52.229	15.999	53.127
First rib	Median	9.635	0.165	0.950	49.290	20.866	50.282
(m = 2)	95% CI	8.412	0.075	0.890	46.794	19.167	47.675
		11.121	0.246	0.990	50.233	22.538	52.177
Pubic symphysis	Median	11.560	0.227	0.950	55.897	24.783	59.965
(m = 3)	95% CI	10.111	0.109	0.890	52.835	22.821	57.551
		13.193	0.349	0.990	58.394	26.674	62.771
Sacroiliac complex	Median	8.575	0.132	0.940	46.654	18.001	51.591
(m = 4)	95% CI	7.379	0.052	0.880	41.134	16.262	49.178
		9.875	0.208	0.990	48.883	19.738	53.900
Acetabulum	Median	8.845	0.134	0.950	42.164	29.244	46.260
(m = 3)	95% CI	7.660	0.056	0.890	40.335	27.221	43.929
		10.053	0.199	0.990	44.378	30.680	49.493
Degenerative traits	Median	6.915	0.080	0.940	32.838	27.965	34.491
(m = 39)	95% CI	6.010	0.004	0.880	30.900	26.413	32.174
		7.896	0.143	0.980	34.478	29.253	36.728
Standard	Median	6.650	0.085	0.940	33.886	11.352	42.638
(m = 16)	95% CI	5.739	0.021	0.880	29.974	10.246	40.344
		7.781	0.147	0.980	37.529	13.042	44.742
Full	Median	5.829	0.064	0.940	29.906	13.305	37.856
(m = 64)	95% CI	4.970	0.002	0.880	26.439	11.633	35.690
		6.686	0.119	0.990	33.383	15.939	40.161

Table 6.5 Monte Carlo cross-validation for models built on different fractions of available skeletal traits. Ensembled deep randomized neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.964	0.120	0.950	30.354	15.851	36.215
(m≈57)	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%	Median	6.026	0.121	0.950	30.498	16.004	36.261
(m≈51)	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%	Median	6.072	0.125	0.950	30.805	16.206	36.454
(m ≈ 44)	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%	Median	6.131	0.125	0.950	30.964	16.352	36.649
(m ≈ 38)	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%	Median	6.237	0.129	0.950	31.479	16.717	36.969
(m ≈ 32)	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%	Median	6.360	0.134	0.950	32.125	17.165	37.429
(m ≈ 25)	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%	Median	6.570	0.140	0.950	33.163	17.933	38.137
(m ≈ 19)	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%	Median	6.951	0.153	0.950	35.263	19.946	39.694
(m ≈ 12)	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%	Median	8.026	0.196	0.950	39.618	26.914	43.025
(m ≈ 6)	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 6.6 Monte Carlo cross-validation for models built on different fractions of available skeletal traits. Ensembled deep randomized neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.964	0.120	0.950	29.623	16.606	37.477
(m≈57)	95% CI	5.136	0.062	0.890	26.347	15.164	35.418
		6.773	0.169	0.990	32.985	18.853	39.732
80%	Median	6.026	0.121	0.950	29.937	16.747	37.579
(m≈51)	95% CI	5.211	0.061	0.890	26.468	14.909	35.469
		6.851	0.172	0.990	32.810	19.445	39.720
70%	Median	6.072	0.125	0.950	30.154	16.960	37.725
(m ≈ 44)	95% CI	5.152	0.062	0.890	26.566	14.751	35.429
		6.924	0.180	0.990	33.499	20.572	40.306
60%	Median	6.131	0.125	0.950	30.371	17.133	37.925
(m ≈ 38)	95% CI	5.316	0.065	0.890	26.661	14.577	35.524
		7.049	0.179	0.990	33.808	21.444	40.635
50%	Median	6.237	0.129	0.950	30.773	17.468	38.242
(m ≈ 32)	95% CI	5.293	0.064	0.890	27.109	14.265	35.521
		7.180	0.179	0.990	34.451	22.972	41.432
40%	Median	6.360	0.134	0.950	31.493	17.827	38.793
(m ≈ 25)	95% CI	5.441	0.074	0.890	27.694	14.436	35.634
		7.380	0.193	0.990	35.170	24.064	42.041
30%	Median	6.570	0.140	0.950	32.382	18.636	39.386
(m ≈ 19)	95% CI	5.565	0.075	0.890	28.479	14.338	35.993
		7.651	0.201	0.990	36.209	26.009	43.067
20%	Median	6.951	0.153	0.950	34.219	20.427	40.971
(m ≈ 12)	95% CI	5.857	0.086	0.890	30.021	14.297	35.884
		8.139	0.218	0.990	38.578	29.418	45.524
10%	Median	8.026	0.196	0.950	38.343	27.355	43.430
(m ≈ 6)	95% CI	6.592	0.119	0.890	33.574	15.518	37.713
		9.683	0.276	0.990	45.094	34.888	50.867

Table 6.7 Monte Carlo cross-validation for models built on different fractions of available skeletal traits. Deep supervised autoencoder neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.906	0.066	0.930	28.737	12.878	34.642
(m≈57)	95% CI	5.056	0.004	0.870	25.455	11.258	32.922
		6.768	0.124	0.980	31.923	15.534	36.385
80%	Median	5.921	0.065	0.930	28.961	12.983	34.837
(m≈51)	95% CI	5.032	0.006	0.870	25.825	11.200	32.875
		6.832	0.122	0.980	32.251	16.048	36.798
70%	Median	5.981	0.065	0.930	29.199	13.447	34.988
(m ≈ 44)	95% CI	5.105	0.002	0.860	26.096	11.196	32.891
		6.967	0.126	0.980	32.592	17.103	37.101
60%	Median	6.059	0.066	0.930	29.610	13.708	35.196
(m ≈ 38)	95% CI	5.243	0.008	0.870	26.362	11.358	33.165
		6.998	0.126	0.980	32.668	18.260	37.237
50%	Median	6.153	0.070	0.930	30.236	14.066	35.555
(m ≈ 32)	95% CI	5.201	0.009	0.870	26.923	11.559	33.371
		7.227	0.129	0.980	33.367	19.050	37.872
40%	Median	6.348	0.074	0.930	30.980	14.823	36.115
(m ≈ 25)	95% CI	5.359	0.009	0.870	27.539	11.697	33.690
		7.380	0.131	0.980	33.968	21.547	38.448
30%	Median	6.550	0.077	0.930	31.857	15.488	36.783
(m ≈ 19)	95% CI	5.475	0.018	0.870	28.123	11.953	34.248
		7.691	0.139	0.980	35.143	23.198	39.735
20%	Median	6.970	0.086	0.930	33.950	17.086	38.311
(m ≈ 12)	95% CI	5.882	0.012	0.880	29.974	12.464	35.193
		8.247	0.159	0.980	38.045	27.193	42.047
10%	Median	8.041	0.113	0.940	38.148	23.368	41.760
(m ≈ 6)	95% CI	6.626	0.032	0.880	33.929	13.528	37.787
		9.781	0.188	0.980	44.379	31.720	47.693

Table 6.8 Monte Carlo cross-validation for models built on different fractions of available skeletal traits. Deep supervised autoencoder neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.906	0.066	0.940	30.174	13.571	38.034
(m≈57)	95% CI	5.056	0.004	0.880	26.674	11.640	35.529
		6.768	0.124	0.990	33.891	16.613	40.472
80%	Median	5.921	0.065	0.940	30.366	13.691	38.148
(m≈51)	95% CI	5.032	0.006	0.890	26.965	11.501	35.706
		6.832	0.122	0.990	34.088	17.057	40.904
70%	Median	5.981	0.065	0.940	30.565	14.098	38.282
(m ≈ 44)	95% CI	5.105	0.002	0.880	26.896	11.609	35.827
		6.967	0.126	0.980	34.089	18.122	41.023
60%	Median	6.059	0.066	0.940	30.901	14.464	38.497
(m ≈ 38)	95% CI	5.243	0.008	0.880	27.273	11.612	35.661
		6.998	0.126	0.980	34.706	19.411	41.373
50%	Median	6.153	0.070	0.940	31.484	14.761	38.751
(m ≈ 32)	95% CI	5.201	0.009	0.880	27.827	11.825	35.839
		7.227	0.129	0.990	35.122	20.019	42.046
40%	Median	6.348	0.074	0.940	32.184	15.465	39.239
(m ≈ 25)	95% CI	5.359	0.009	0.880	28.529	12.048	36.218
		7.380	0.131	0.990	35.945	22.740	42.593
30%	Median	6.550	0.077	0.940	33.077	16.225	39.944
(m ≈ 19)	95% CI	5.475	0.018	0.880	28.895	12.083	36.288
		7.691	0.139	0.980	36.873	24.086	43.802
20%	Median	6.970	0.086	0.940	35.072	17.633	41.193
(m ≈ 12)	95% CI	5.882	0.012	0.890	30.684	12.528	36.698
		8.247	0.159	0.980	39.305	28.050	45.879
10%	Median	8.041	0.113	0.940	39.414	24.202	44.367
(m ≈ 6)	95% CI	6.626	0.032	0.890	33.942	13.187	38.642
		9.781	0.188	0.990	46.173	32.477	51.486

Experiment B (Table 6.5 to Table 6.8) exhibited from a technical perspective a very similar results with relative absolute differences across model parameterizations being minimal for the median values of the metrics computed. From an anthropological perspective the striking result of this experiment is that comparable performances to the full multi-trait model (m=64) were obtained using different proportions of traits selected at random. Virtually every combination of random traits resulted in models with comparable or better performance than the models built of specific anatomic regions as traditionally encounter in macroscopic age estimation methods. This finding supports the value of multifactorial age estimation over methods the focus solely on a single anatomical structure.

An important remark to make regarding the results based of the two computational experiments is that the analytical leave-one out cross-validation (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. Detailed tabular results are available in appendix C.

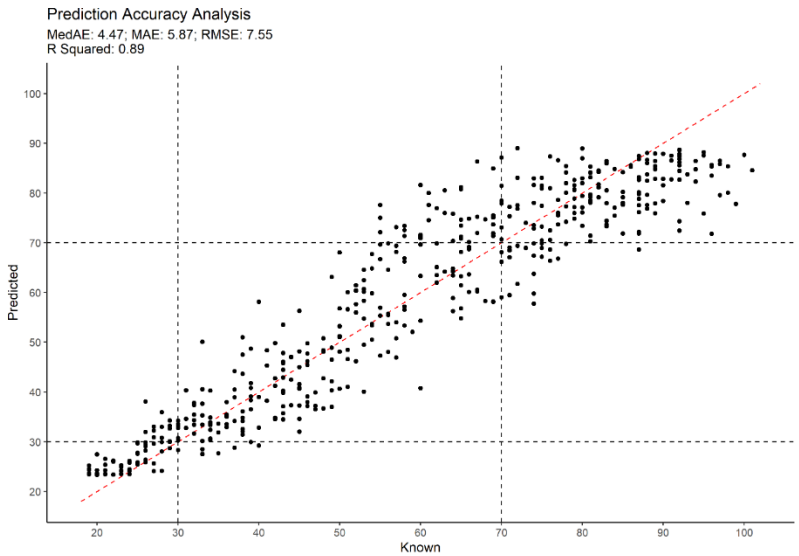

Figure 6.5 Known vs. predicted age-at-death using a full set of traits (LOOCV, n = 500).

The maximal potential accuracy of this approach can be visualized in Figure 6.5, where a scatter plot of known vs. predicted age-at-death is depicted. From this figure, one can infer that the predictions obtained using this 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 was also possible to visualize, Figure 6.1, that a deep RANN model using multiple traits produces minimally biased estimates. Multifactorial models also show a systematical reduction in prediction bias when compared to models based only on a specific anatomical structure.

Regarding the validity of the models trained in the 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 $P(\alpha) \sim 0.95$).

6.5|Discussion

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. 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. Results from 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 [141, 142]; 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 [144, 145]. 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 18-46 years. 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). Anthropologists and bioarcheologists have made hazy and sometimes arbitrary distinctions between *normal* aging, disease, and activity based on their education, experience, and research objectives. Most studies on joint and musculoskeletal degenerative traits have focused on the premise of its relationship with activity patterns or disease. Age-at-death is one of the most significant, or even the only factor with statistical significance, in the expression of such skeletal traits, according to recent and systematic studies conducted on identified skeletal collections [143–145, 178–182]. The results from this thesis substantiate their importance in age estimation of the elderly.

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 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 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, reinforcing the hypothesis 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 was the main reason why it was opted for a randomized evaluation of smaller traits sets.

From a practitioner perspective, correlation analysis (Chapter 4) 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, 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. [154], who also investigated the cranial sutures in the context of a multifactorial approach. Despite its limited value for age estimation, according to Teixeira and Cunha [270], and Lourenço and Cunha[271], cranial sutures and other cranial traits can act as secondary age indicators never to be used alone.

To maximize the potential of the framework proposed, 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. [159]. 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). 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, it was 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 this 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 [125].

The present work provides a solution to the problem of multifactorial age estimation based on the macroscopic analysis of the skeleton. A multifactorial perspective is systematically noted as being the most accurate way to achieve age estimation in adults, but is obtained through a plethora of procedures and heuristics that are often subjective and lack a clear statistical or computational rationale [41, 43].

As noted by Ritz-Timme et al. [43], 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).

These limitations are exacerbated by the fact that no other study published in the literature have provided objective performance analysis of adult skeletal age estimation using such a vast and diverse array of macroscopic traits based on a single reference dataset,

despite that Milner and collaborators [155, 272] are working on similar approach to the one here presented. Nonetheless, an analysis of the most recent and comprehensive validation studies clearly demonstrates that the multifactorial approach, offers improved accuracy (MAE $<$ 8 years) in relation to other skeletal age estimation methods [154, 273–275]. 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 [31, 51–53, 61, 62, 81]. 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 [276]. In this study, was used a resampling strategy based on Monte Carlo cross-validation for fair model assessment. Analytical and computationally efficient leave-one-out cross-validation strategy was used to set the regularization parameter of the networks developed in experiments A and B.

Results support that a regression-based framework produces accurate age estimation in adult individuals. Prediction intervals can be estimated with ease and computational efficiency. Probabilistic approaches [54, 56, 58, 62, 218] 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 pursued here. Recent contributions suggest that Bayesian approaches do not radically improve age-at-death estimation or outperform regression-based approaches [204, 277].

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

7|A novel software for adult skeletal age estimation

As previously demonstrated a predictive modelling strategy supported by a regression-based framework using deep randomized neural networks and uncertainty modelling techniques is an effective method of obtaining accurate, unbiased, valid, and efficient age-at-death estimation models. The current work provides a strong and favorable argument for multifactorial age estimation over traditional methods that target specific anatomic structures. However, fields such as forensic anthropology demand that the results of research demonstrate usefulness and practical application. The main goal of research should be to implement and operationalize new knowledge in the form of guidelines, standard operating procedures, or new tools and methods. The purpose of this chapter is to provide a general introduction to the software known as DRNNAGE (Deep Random Neural Networks for Adult Skeletal Age-at-Death Estimation) [198]. This novel piece of software implements and integrates in a straightforward and user-friendly graphical user interface, the capability to estimate age-at-death from adult skeletal remains on a case-by-case basis using the described macroscopic procedures and computational models.

7.1|License

DRNNAGE and its underlying components (i.e., additional software packages) are released as free and open-source software under the version 3 or later of the GNU General Public License (GPLv3).

The adoption of open-source and free software philosophy is a two-fold by-product of both technoscientific and ethical considerations. Some organizations and research groups charge fees for simple computations and/or routines that may be performed on extensible, open-source platforms, hence rendering the expense unnecessary in the context of biological profile estimation. As technology progresses and free open-source software become more

popular and successful, they may displace commercial closed-source tools and become more prevalent [219]. It is important to note that commercialized closed-source technologies may be unreliable, offering little advantage over open-source equivalents despite their high cost. While an open-source philosophy does not preclude faults, it can provide a platform and framework for rapid issue resolution in the sense of community engagement and scientific progress.

In forensic anthropology, a lack of adequate resources can hinder or make software testing incomplete. If the code is faulty, insufficient testing may damage the utility and reputation of a tool. In the ethos of science, noncommercial open-source software allows unrestricted access to the scientific community while avoiding any ethical gray areas regarding financial benefit from the analysis of decedents and/or crime victims [278].

7.2|Availability

DRNNAGE is primarily intended to be used and accessed as a web application at **https://osteomics.com/DRNNAGE/**. It integrates the toolbox provided by Osteomics (**https://osteomics.com**) a web-based platform that make freely available decision support systems for biological and forensic anthropology. The online applications include regression and classification-based computational solutions for sex, age-at-death, biogeographic ancestry, and body parameters (i.e., stature and skeletal mass) estimation. The applications have a user-friendly interface, detailed documentation (redirecting to relevant literature), data exploration (including graphical analysis), and regression and classification outputs. In addition, some of the available web applications provide a set of features with pedagogical significance in the fields of human skeletal anatomy and graphical and statistical analysis [279].

Source code of the software and its constituting components are under version control using the Git system [280] and the GitHub service [281] and archiving using the open science platform Zenodo [220]. Table 7.1 lists the repositories hyperlinks for the source

code of the packages created and integrated in the DRNNAGE software.

Table 7.1 Repositories of source code for software packages developed during this thesis.

Package	Title	Repository
DRNNAGE	Deep random neural networks for adult skeletal age-at-death estimation	https://github.com/dsnavega/DRNNAGE
rwnnet	An R/C++ Implementation of Random Weights Neural Networks	https://github.com/dsnavega/rwnnet
rumr	Regression Uncertainty Modeler in R	https://github.com/dsnavega/rumr
rmar	Regression Model Assessment in R	https://github.com/dsnavega/rmar
lsmr	Linear Surrogate Model for Regression in R	https://github.com/dsnavega/lsmr

7.3|Development

DRNNAGE was designed with web deployment as its primary point of access. Nonetheless, it can also be deployed offline as a package for the R programming language and statistical computing environment [282]. R is a high-level programming language and environment for statistical computing and data visualization created by John Chambers and colleagues at Bell Laboratories (previously AT&T, now Lucent Technologies). R is highly extendable and supports a wide range of statistical (linear and nonlinear modeling, traditional statistical tests, time-series analysis, classification, clustering, etc.), machine learning, and data visualization techniques. R is distributed as an open-source free software under the provisions of the Free Software Foundation's GNU General Public License. It compiles and operates on a wide range of UNIX and comparable systems (including FreeBSD and Linux), as well as Windows and MacOS. Higher-level languages allow for faster prototyping and tool development. R can integrate with C++, C, and Python to improve processing speed and automate tasks. It also connects with web technologies, allowing for the creation and deployment of data and computationally intensive online

apps and dashboards [282]. R is a prominent environment for statistical and morphometric analysis in forensic and biological anthropology, as well as software development and analytical tools used by practitioners in these fields. R and its ecosystem are the primary technologies used by 56% of forensic anthropological software tools [278].

DRNNAGE interface and web deployment was built and made possible due to the *shiny* package [283] and *shiny webserver* [284]. *shiny* is a R package that makes it simple to create interactive web applications directly in R. Developed applications can be hosted on a website or embed in R Markdown documents. *shiny* applications can also be extended with CSS themes, HTML widgets, and JavaScript actions. *shiny server* provides a platform for hosting multiple *shiny* applications on a single server, with each application having its own URL and port. It has support to non-WebSocket enabled browsers such as Internet Explorer 10 and is licensed under AGPLv3. Figure 7.1 shows the landing page for the DRNNAGE software as rendered in *Microsoft Edge* web browser.

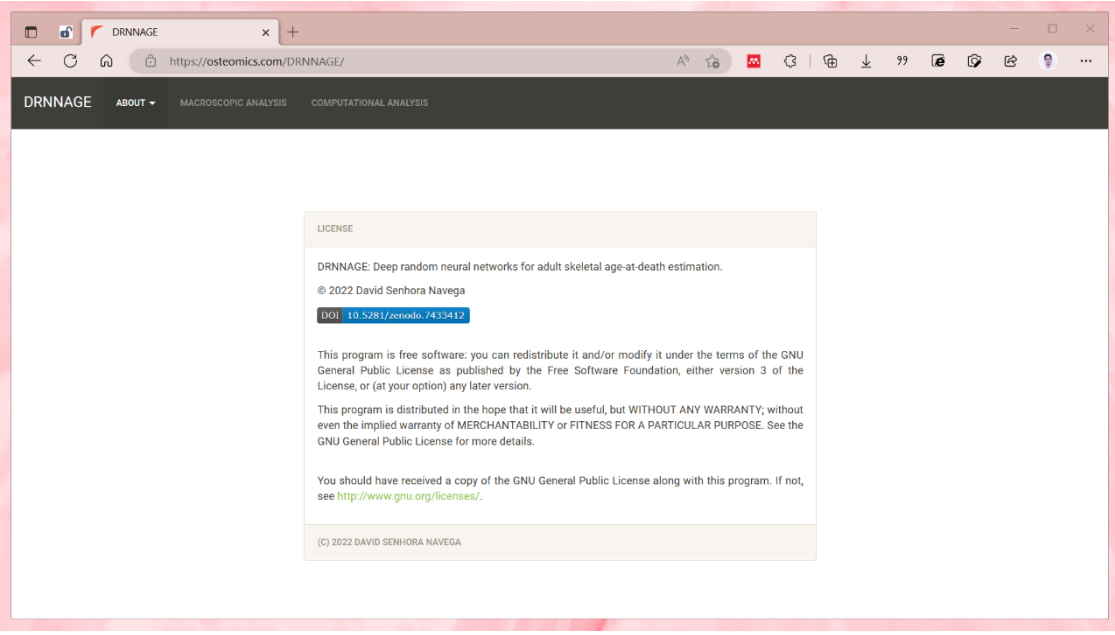

Figure 7.1 DRNNAGE landing page (https://osteomics.com/DRNNAGE)

R was used as the main programming language and ecosystem in the development of DRNNAGE and all the subcomponents that enable its functionalities. A notable exception to this is the *rwnnet* R package (Table 7.1) which has its core written in C++. Integration of R and C++ was accomplished by using the *Rcpp* and *RcppArmadillo* packages by Eddelbuettel [285–288]. *RcppArmadillo* enables access to Armadillo, a template-based C++ high-performance library for linear algebra based computation [278, 289]. Throughout Chapter 5 it has been demonstrated that deep randomized neural networks, as used in this work, can be fully represented, and operationalized via matrix computations. A C++ implementation combined with the mathematical formulation used in this thesis for training, optimization of the regularization parameter in the output layer, enabled fast and efficient neural network model training and validation.

7.4|Functionalities

The main function of this software is to enable flexible and case-by-case age-at-death estimation of adult skeletal human remains from the available and scorable age-related traits. The software can create new neural network models for any combination of skeletal traits given that at least two traits are available as input. It integrates and operationalizes the key contributions of this work: the macroscopic technique developed to assess age-related traits (Chapter 3) and computational approach used to generate estimates from the observed skeletal traits (Chapter 5) using as a reference the curated dataset (Chapter 4). Model assessment metrics are also computed to evaluate the model performance (Chapter 6). The interface is characterized by two major components: the macroscopic and the computational analysis tabs whose main functionalities are described in more detail in the following sections.

7.4.1|Input

On the macroscopic analysis tab, data is entered using intuitive radio button components. All values are initially set to NA, which stands for non-availability or not scored (NA is also the internal representation of R for missing values). For bilateral traits, both sides can be used as input, but the software will only use the right side if the left side is missing (NA). The textual description of each trait score is embedded in sub-tabs for each skeletal trait so that practitioners have self-contained access to the scoring system. Figure 7.2 and Figure 7.3 depict data input interface and traits description on the macroscopic analysis tab.

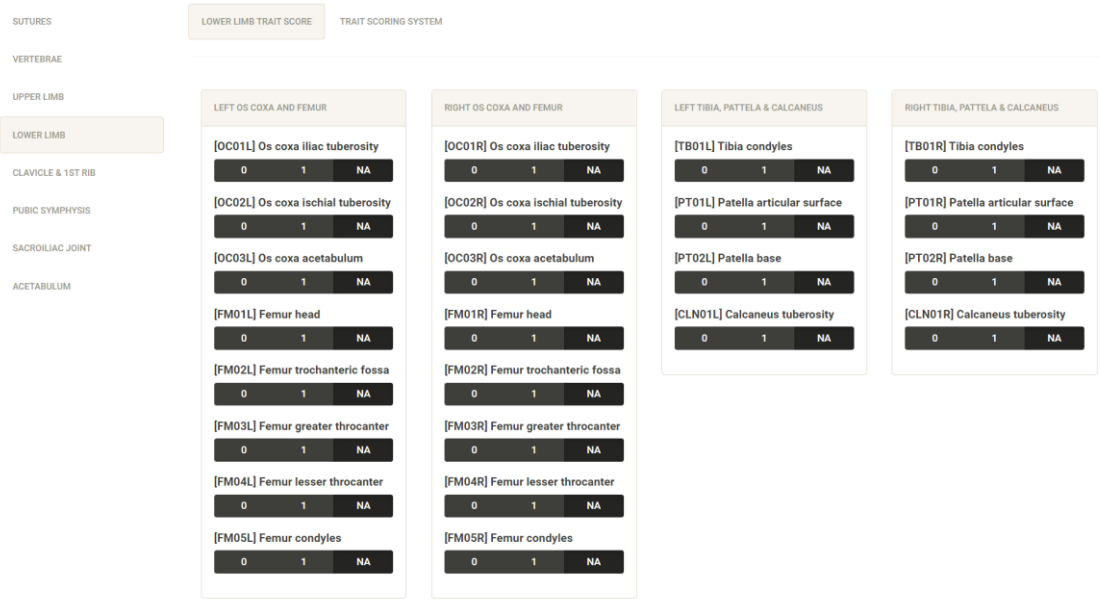

Figure 7.2 Data input via radio buttons on the macroscopic analysis tab.

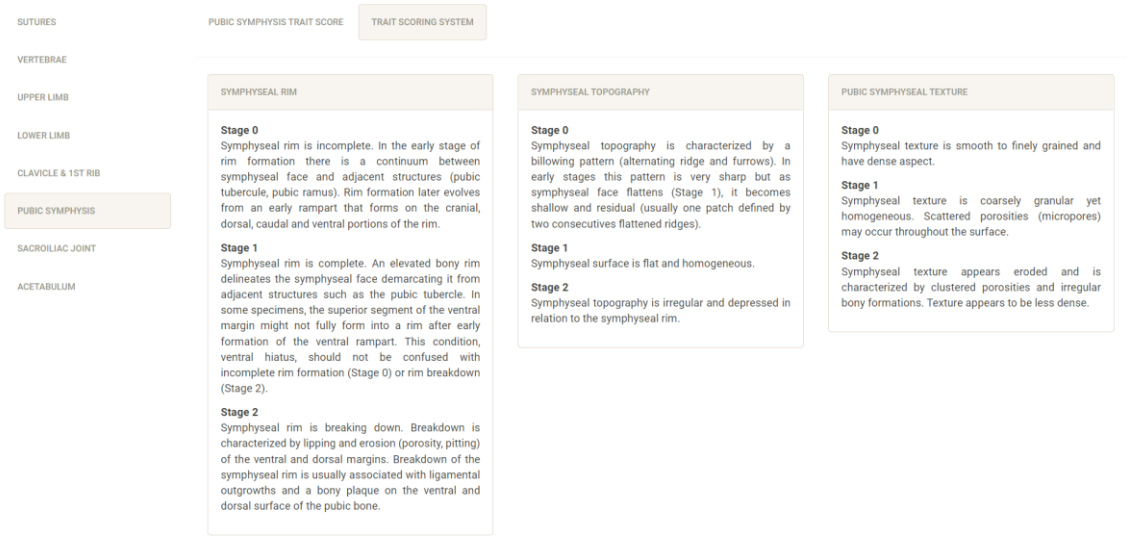

Figure 7.3 Textual trait scoring system description sub-tab example.

7.4.2|Estimate

As previously stated, the primary goal of this software is to obtain age-at-death estimates on a case-by-case basis. This is done through the computational analysis tab, which has three main sub-component tabs: estimate, explain, and asses. The **estimate tab** is the main output component of the application. If valid data is available, pressing the analyze button will output the estimated (or predicted) age-at-death for a given skeletal pattern. The conditional mean age-at-death estimate provided to the user is based on the skeletal traits observed. The average of the two uncertainty modeling strategies proposed in this work yields the final predictive interval. A visual representation of the uncertainty associated with an individual estimation is obtained by representing a truncated gaussian distribution parameterized with the conditional mean and variance as described in Chapter 5. A report will also be generated as an HTML document by the software, which can be printed or saved as a portable document file (.pdf) using a modern web browser such as Google Chrome, Microsoft Edge, or Mozilla Firefox. When you press the report button, a report is generated and downloaded. This report summarizes all the information from the

sub-tabs of the computational analysis and parameterization used. Figure 7.4 and Figure 7.5 provide two examples of estimates obtained for a young and an elderly individual.

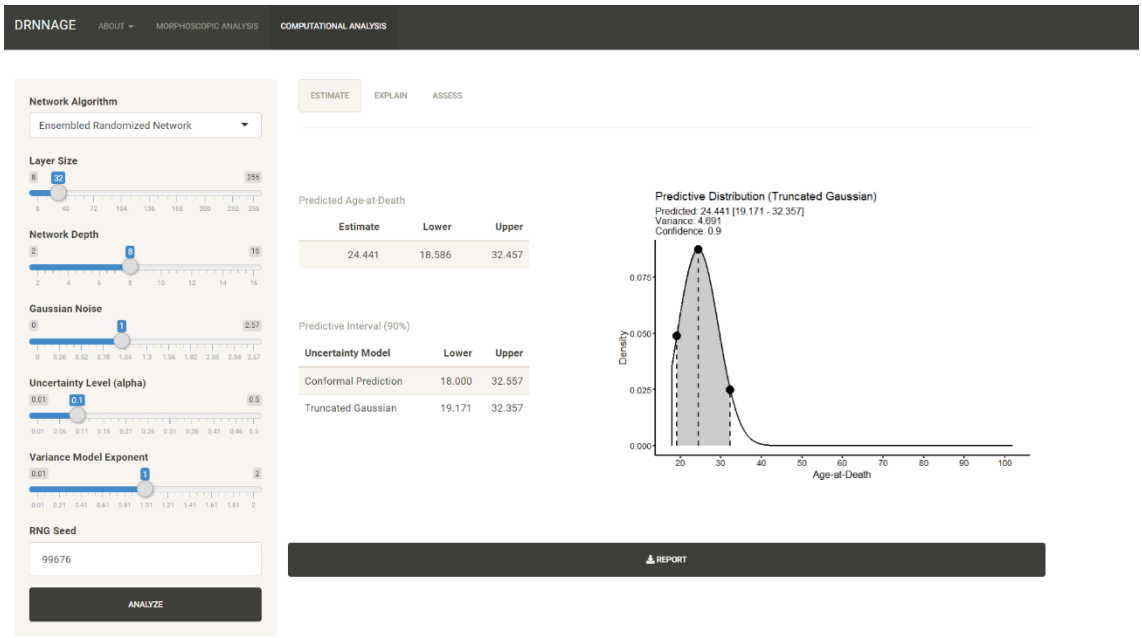

Figure 7.4 Estimate obtained by DRNNAGE on a young individual (25 years old).

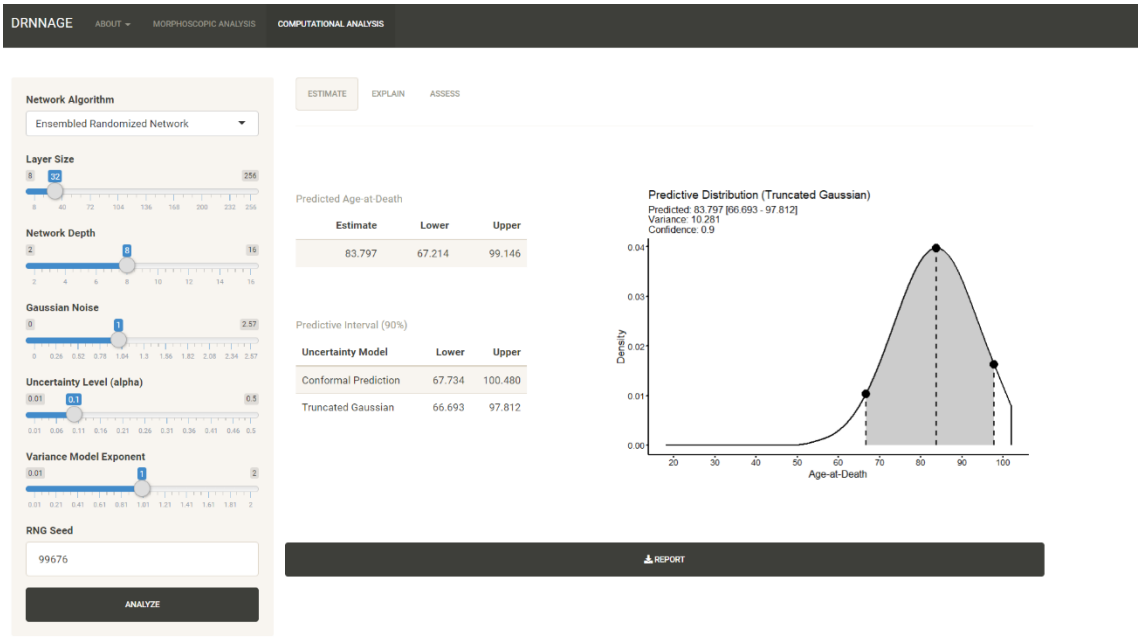

Figure 7.5 Estimate obtained by DRNNAGE on an elderly individual (88 years old).

7.4.3|Explain

Predictions or estimates made by artificial neural networks models often shows remarkable accuracy yet these models, and other machine learning techniques as well, are frequently regard them as *black boxes*. Humans find it difficult to gain insights into decision making involved in such predictive algorithms. Understanding decision making in extremely sensitive domains is especially important. The decision-making behind the machine learning decision support systems must be more transparent, accountable, and understandable to humans [290]. 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 with new methods and approaches for the interpretation being are published at staggering rate [290–295]. It is not the objective of this section to delve deeper on what constitutes a research problem on its own, but provide the necessary background to understand the rationale and approach implemented in DRNNAGE regarding the issue interpretability and explainability of the models and estimates computed by the software.

A neural network model built with the approach proposed in this work can be seen as a regularized linear model that operates on the feature space composed of the non-linear input extracted by the hidden layers and the initial skeletal traits. This approach's algorithmic transparency is comparable to a linear model, as shown in Chapter 5. Unlike a linear model, the mathematical features and randomized nature of hidden layers make the number of coefficients too large for human comprehension, even though the estimate is defined by linear combination. To improve the interpretability and explainability of DRNNAGE models and estimates, a global surrogate model approach [291] using a linear model and data transformation was implemented. This approach exploits the intuitive and additive nature of linear models and their relationship to an artificial neural network's output layer in the context of numeric prediction. In the global surrogate model approach a simpler and interpretable model such a linear model is used to approximate the estimates obtained from a more complex model. In the case of linear model, an interpretable

approximation of the network estimate using a surrogate model is written as

$$f^*(z) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p \tag{7.1}$$

where the linear coefficients, $\boldsymbol{\beta}$, are in this context obtained by regress the (cross-validated) estimates obtained from the neural network model, $\hat{y} = f(x, \theta)$, on the decorrelated version of the network initial inputs, $\boldsymbol{Z}$.

To decorrelate the initial input, the data is centered and scaled before being transformed using a linear basis that enforces orthogonality among predictors while preserving the maximum relationship with the original input. After this transformation, also known as the Mahalanobis transform or sphering [296, 297], the covariance matrix of the predictors, $\boldsymbol{Z}$, is the identity matrix. The use of this data transformation removes multicollinearity among the inputs of the interpretable model — enforcing no correlation among predictors, an assumption of multiple linear regression. Given that the transformed data is also centered around the origin (0) and a unit variance (1) the intercept, β_0, is the mean value of the estimates of neural network, $f(x, \theta)$, and a natural way to measure the contribution or impact of a particular trait, p, on the estimate as approximated by the surrogate model is given by $\beta_i z_i$.

An important by-product of the Mahalanobis-decorrelated inputs used in the interpretable models is that the linear correlation coefficient is equivalent to what Zuber and Strimmer [296] call CAR scores (Correlation-Adjusted (marginal) coRrelation), defined as the marginal correlations adjusted for correlation among explanatory variables, a natural variable importance criterion. The square of the CAR scores and its sum leads to a simple additive decomposition of the proportion of explained variance.

When deploying a machine learning system is important to guarantee what Murdoch et al. [293] refers as predictive accuracy, descriptive accuracy and relevance. In this context, predictive accuracy is accomplished by the neural network models which capture the underlying relationship between the skeletal traits and age-at-death and can

predict or estimate with remarkable levels of accuracy. Murdoch et al. [293] defines descriptive accuracy as the degree to which an interpretation method objectively captures the relationships learned by machine-learning models. In the **explain tab**, to ensure the fidelity of the surrogate model to the neural network, several metrics (MAE, RMSE, R^2, Bias) are reported to assess the quality of the surrogate model (Figure 7.6). The addition of a post hoc interpretation method enhances the relevance of this tools which already provides a significant solution to the vexing problem of age estimation in the context of forensic human identification.

This implementation raises the question: why not directly predict age-at-death with a linear model? The neural network approach pursued in this study, can be viewed as an extension of a linear model with nonlinearities via activation functions. The linear surrogate model provides interpretability in manner that is familiar to forensic experts. This integrative approach paves the way for improved decision support systems, as opposed to pointless debates over algorithmic preferences.

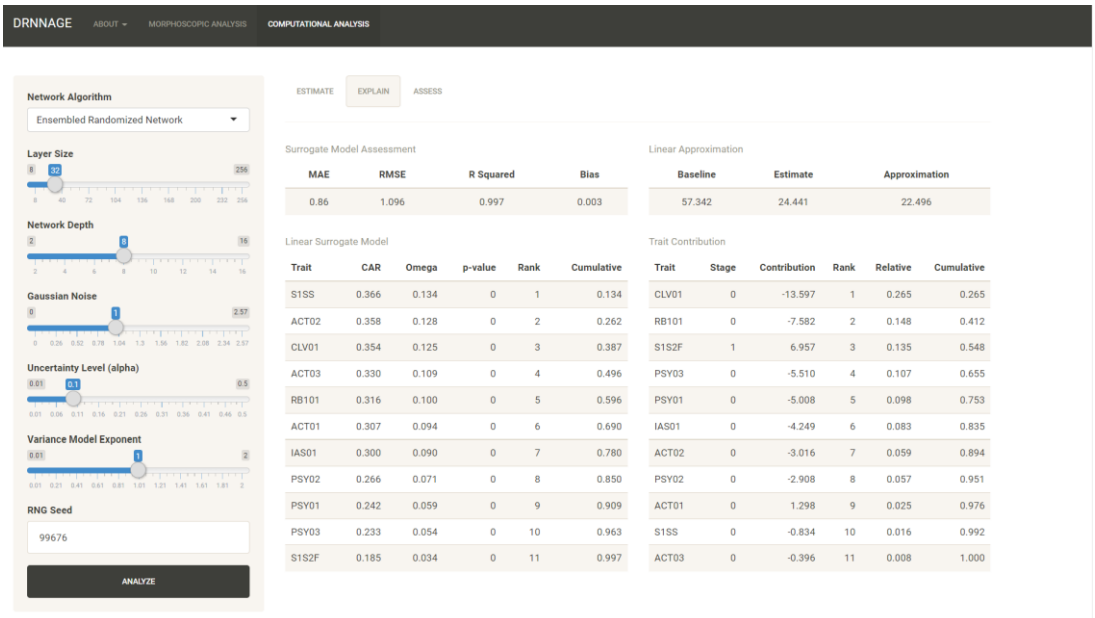

Figure 7.6 Interpretability and explainability via linear surrogate model.

7.4.4|Assess

Due to the flexible nature of the software, it is important for the end user to have feedback on the predictive accuracy of the model being used for age estimation on case-by-case basis. Along with on-the-fly model fitting, the software also performs implicit model validation and assessment. As detailed in Chapter 5 the formulation of neural networks used in the work allows for analytical leave-one-out estimates to be obtained, which can then be used as robust baseline for model assessment without resorting to an external dataset or more computationally intensive cross-validation techniques. This feature enables fast model validation, a critical feature for a software with web deployment as its main access-point. The **assess tab** provides both numeric and graphical output to enable model evaluation (Figure 7.7).

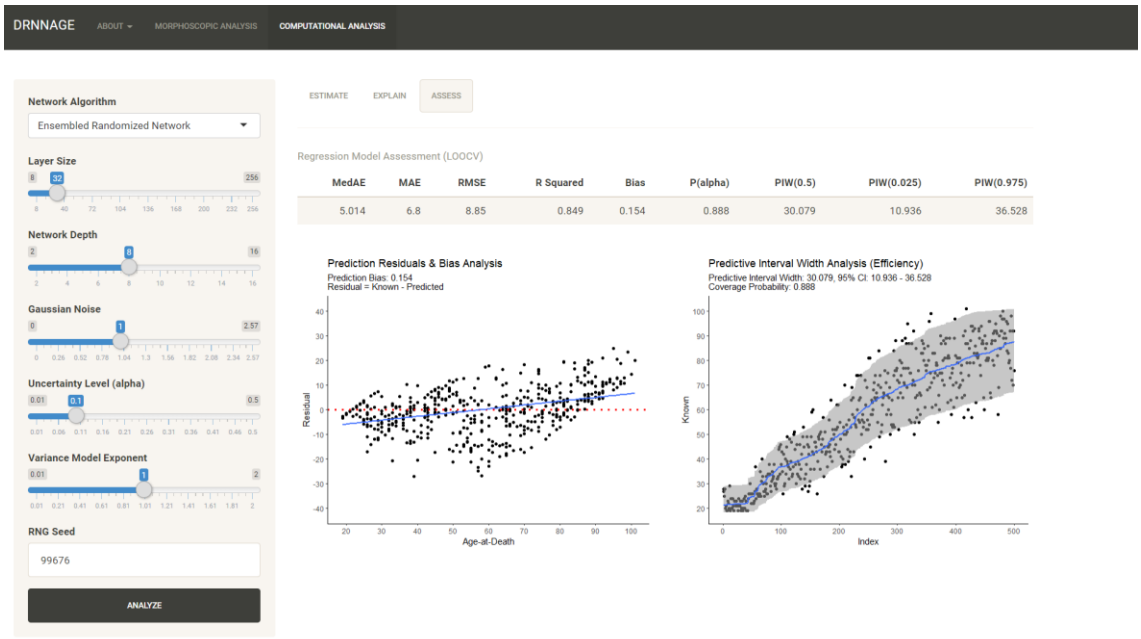

Figure 7.7 Model assessment and evaluation output.

7.5|Parameterization

Some degree of parameterization or fine-tuning is achievable in the software. By default the software fits a ensembled randomized network (network algorithm select box) composed with of 8 layers (network depth slider) with 32 ReLU units each (layer size slider), all components of the network as disturbed with a moderate amount of gaussian noise (slider, set to 1) for robust training, the variance model regressor operates in a linear fashion (slider, set 1) and predictive intervals are created for 90% confidence level (slider, 1-alpha value). This are sensible values chosen based on domain knowledge and extensive testing during software development.

The usage of skip layers, re-use of the initial input along the layers of the network path, render the models highly stable and comparable across parameterizations both in accuracy and computational performance. Variations are mostly due to the intrinsic nature of this approach, randomization (controlled by the seed numeric box). A notable exception to this is the use of the ensembled supervised autoencoder algorithm which shows an autocorrective behavior in relation to predictive bias. This algorithm tends to produce less biased models which can be important at the extremes of the adult lifespan, more accurate estimate for very young or elderly individuals. In the current state more important than the parameterization used, is to keep record of the parameters. The report generated by the software tracks the parameters (Figure 7.8).

Parameter	Value
Network	Ensembled Randomized Network
Width	32
Depth	8
Gaussian Noise	1
Alpha	0.1
Exponent	1
Seed	99676

Figure 7.8 Parameterization as reported in DRNNAGE.

[*This page has been intentionally left blank*]

8|Conclusions and future work

The main objective of this work was to investigate the fundamental issue of age-at-death estimation in the 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 can produce 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 [41]. 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 [59, 62, 81]. More generic procedures for multifactorial analysis have also been proposed, but with poor adoption in forensic casework because they require seriation or advanced mathematical knowledge to be put into action [153, 156–158].

This thesis demonstrated that a multifactorial approach to skeletal age estimation improves accuracy and precision over single anatomic regions, as established in traditional adult skeletal aging methods. Multifactorial models introduce a two-to-six-fold reduction in the mean absolute error and prediction bias compared to standard models.

The 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 this case, random weights and a non-linearity (ReLU function) as a mechanism to combine multiple traits and obtain a more accurate latent representation of age-related morphology, 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.

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 because the presented results, as solid as they are, reflect only *in silico* performance assess via cross-validation.

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, it was inspired by significant work that was previously developed on this topic [16,19,24,35,140].

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

The major research output of this thesis is the open-source software DRNNAGE. Randomization of the hidden layers, combined with an efficient C++ implementation of the models developed, allowed the construction of a 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.

This software was built to operationalize the age estimation procedure described in this thesis, 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, rather than the technical or

methodological implementation. The software provides intuitive interface and is freely distributed under an open-source license with web-based access-point for multiplatform support tackling this way issues of accessibility and usability, major obstacles to casework usage of computationally advanced methods [151].

One last aspect that deserves attention 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.

Future work in age estimation should seek to create truly multifactorial methods, integrating multiple modalities of data collection and disciplines. Deep artificial neural networks or deep learning excels in problems involving vision tasks such as the image analysis involved in bone and dental histological or radiological methods with potential for automated feature extraction and age estimation. Explainable machine learning was briefly addressed in this thesis through the inclusion of an explanation system based on global surrogate model approach in the developed age estimation software, this topic should be further explored.

[*This page has been intentionally left blank*]

References

1. Dirkmaat DC, Cabo LL, Ousley SD, Symes SA (2008) New Perspectives in Forensic Anthropology. Am J Phys Anthropol 137:33–52. https://doi.org/10.1002/ajpa.20948
2. Dirkmaat DC, Cabo LL (2015) Embracing the New Paradigm. In: Dirkmaat DC (ed) A Companion to Forensic Anthropology, 1st Ed. Blackwell Publishing Ltd, pp 1–40
3. Işcan MY (1988) Rise of forensic anthropology. Am J Phys Anthropol 31:203–230
4. Council NR (2009) Strengthening Forensic Science in the United States. National Academies Press, Washington, D.C.
5. Christensen AM, Crowder CM (2009) Evidentiary standards for forensic anthropology. J Forensic Sci 54:1211–1216. https://doi.org/10.1111/j.1556-4029.2009.01176.x
6. Grivas CR, Komar DA (2008) Kumho, Daubert, and the nature of scientific inquiry: Implications for forensic anthropology. J Forensic Sci 53:771–776. https://doi.org/10.1111/j.1556-4029.2008.00771.x
7. Iscan MY, Steyn M (2013) The human skeleton in forensic medicine. Charles C Thomas Publisher
8. Ashiqur Rahman S, Giacobbi P, Pyles L, et al (2021) Deep learning for biological age estimation. Brief Bioinform 22:1767–1781. https://doi.org/10.1093/bib/bbaa021
9. Nakamura E, Miyao K, Ozeki T (1988) Assessment of biological age by principal component analysis. Mech Ageing Dev 46:1–18. https://doi.org/10.1016/0047-6374(88)90109-1
10. Klemera P, Doubal S (2006) A new approach to the concept and computation of biological age. Mech Ageing Dev 127:240–248. https://doi.org/10.1016/j.mad.2005.10.004
11. Jia L, Zhang W, Chen X (2017) Common methods of biological age estimation. Clin Interv Aging 12:759–772. https://doi.org/10.2147/CIA.S134921

12. Jackson SHD, Weale MR, Weale RA (2003) Biological age - What is it and can it be measured? Arch Gerontol Geriatr 36:103–115. https://doi.org/10.1016/S0167-4943(02)00060-2

13. Cunha E, Baccino E, Martrille L, et al (2009) The problem of aging human remains and living individuals: A review. Forensic Sci Int 193:1–13. https://doi.org/10.1016/j.forsciint.2009.09.008

14. Black S, Aggrawal A, Payne-James J (2011) Age estimation in the living: the practitioner's guide. John Wiley \& Sons

15. Stull KE, Corron LK, Price MH (2021) Subadult age estimation variables: Exploring their varying roles across ontogeny. In: Algee-Hewitt BFB, Kim J (eds) Remodeling Forensic Skeletal Age. Academic Press, pp 49–73

16. Arking R (2006) Biology of aging: observations and principles, 3rd ed. Oxford University Press

17. Mays S (2015) The effect of factors other than age upon skeletal age indicators in the adult. Ann Hum Biol 42:330–339. https://doi.org/10.3109/03014460.2015.1044470

18. Merritt CE (2015) The influence of body size on adult skeletal age estimation methods. Am J Phys Anthropol 156:35–57. https://doi.org/10.1002/ajpa.22626

19. Wescott DJ, Drew JL (2015) Effect of obesity on the reliability of age-at-death indicators of the pelvis. Am J Phys Anthropol 156:595–605. https://doi.org/10.1002/ajpa.22674

20. Ubelaker DH, Khosrowshahi H (2019) Estimation of age in forensic anthropology: historical perspective and recent methodological advances. Forensic Sci Res 4:1–9. https://doi.org/10.1080/20961790.2018.1549711

21. Ubelaker DH, Longeway A (2019) Skeletal age estimation of the living and the dead: the evolution of methodology. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 29–40

22. C. Zapico S, DeGaglia CM, Adserias-Garriga J (2019) Age estimation based on chemical approaches. In: Adserias-Garriga J (ed) Age Estimation: A

Multidisciplinary Approach. Academic Press, pp 199–211

23. C. Zapico S, Thomas C, Zoppis S (2019) Age estimation based on molecular biology approaches. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 213–223
24. C. Zapico S, Stone-Gordon R, Adserias-Garriga J (2019) The evolution of methodology in biochemical age estimation. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 189–197
25. Dias HC, Cordeiro C, Pereira J, et al (2020) DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. Forensic Sci Int 311:110267. https://doi.org/10.1016/J.FORSCIINT.2020.110267
26. Correia Dias H, Corte-Real F, Cunha E, Manco L (2020) DNA methylation age estimation from human bone and teeth. https://doi.org/101080/0045061820201805011 54:163–176. https://doi.org/10.1080/00450618.2020.1805011
27. Correia Dias H, Cordeiro C, Corte Real F, et al (2020) Age Estimation Based on DNA Methylation Using Blood Samples From Deceased Individuals. J Forensic Sci 65:465–470. https://doi.org/10.1111/1556-4029.14185
28. Correia Dias H, Manco L, Corte Real F, Cunha E (2021) A blood–bone–tooth model for age prediction in forensic contexts. Biology (Basel) 10:1312. https://doi.org/10.3390/BIOLOGY10121312/S1
29. Manco L, Dias HC (2022) DNA methylation analysis of ELOVL2 gene using droplet digital PCR for age estimation purposes. Forensic Sci Int 333:111206. https://doi.org/10.1016/J.FORSCIINT.2022.111206
30. Dominguez VM, Mavroudas S (2019) Bone histology for skeletal age-at-death estimation. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 145–159
31. Curate F, Navega D, Cunha E, Coelho J d'Oliveira (2022) DXAGE 2.0 - adult age at death estimation using bone loss in the proximal femur and the second metacarpal. Int J Legal Med 136:1483–1494. https://doi.org/10.1007/S00414-022-

02840-Y

32. Navega D, d'Oliveira Coelho J, Cunha E, Curate F (2018) DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. J Forensic Sci 63:497–503. https://doi.org/10.1111/1556-4029.13582
33. Castillo RF, López Ruiz M del C (2011) Assessment of age and sex by means of DXA bone densitometry: Application in forensic anthropology. Forensic Sci Int 209:53–58. https://doi.org/10.1016/j.forsciint.2010.12.008
34. Bethard JD, Berger JM, Maiers J, Ross AH (2018) Bone Mineral Density Adult Age Estimation in Forensic Anthropology: A Test of the DXAGE Application. J Forensic Sci 3–6. https://doi.org/10.1111/1556-4029.13987
35. Wade A, Nelson A, Garvin G, Holdsworth DW (2011) Preliminary radiological assessment of age-related change in the trabecular structure of the human os pubis. J Forensic Sci 56:312–319. https://doi.org/10.1111/j.1556-4029.2010.01643.x
36. Cameriere R, Cunha E, Sassaroli E, et al (2009) Age estimation by pulp/tooth area ratio in canines: Study of a Portuguese sample to test Cameriere's method. Forensic Sci Int 193:128.e1-128.e6. https://doi.org/10.1016/j.forsciint.2009.09.011
37. Cameriere R, Ferrante L, Belcastro MG, et al (2007) Age estimation by pulp/tooth ratio in canines by peri-apical X-rays. J Forensic Sci 52:166–170. https://doi.org/10.1111/j.1556-4029.2006.00336.x
38. Lamendin H, Baccino E, Humbert JF, et al (1992) A simple technique for age estimation in adult corpses: the two criteria dental method. J Forensic Sci 37:13327J. https://doi.org/10.1520/jfs13327j
39. de las Heras SM (2019) Dental age estimation in adults. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 125–142
40. Bertrand B, Oliveira-Santos I, Cunha E (2019) Cementochronology: a validated but disregarded method for age at death estimation. In: Adserias-Garriga J (ed) Age Estimation: A Multidisciplinary Approach. Academic Press, pp 169–186
41. Garvin HM, Passalacqua N V. (2012) Current Practices by Forensic Anthropologists

in Adult Skeletal Age Estimation. J Forensic Sci 57:427–433. https://doi.org/10.1111/j.1556-4029.2011.01979.x

42. Campanacho V, Chamberlain AT, Nystrom P, Cunha E (2020) Degenerative variance on age-related traits from pelvic bone articulations and its implication for age estimation. Anthropol Anz 77:243–258. https://doi.org/10.1127/ANTHRANZ/2020/1184

43. Ritz-Timme S, Cattaneo C, Collins MJ, et al (2000) Age estimation: The state of the art in relation to the specific demands of forensic practise. Int J Legal Med 113:129–136. https://doi.org/10.1007/s004140050283

44. Ferrante L, Cameriere R (2009) Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. Int J Legal Med 123:277–283. https://doi.org/10.1007/s00414-009-0349-4

45. Franklin D (2010) Forensic age estimation in human skeletal remains: Current concepts and future directions. Leg Med 12:1–7. https://doi.org/10.1016/j.legalmed.2009.09.001

46. Rösing FW, Graw M, Marré B, et al (2007) Recommendations for the forensic diagnosis of sex and age from skeletons. HOMO- J Comp Hum Biol 58:75–89. https://doi.org/10.1016/j.jchb.2005.07.002

47. Kimmerle EH, Prince DA, Berg GE (2008) Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth. J Forensic Sci 53:594–600. https://doi.org/10.1111/j.1556-4029.2008.00715.x

48. Martrille L, Ubelaker DH, Cattaneo C, et al (2007) Comparison of four skeletal methods for the estimation of age at death on white and black adults. J Forensic Sci 52:302–307. https://doi.org/10.1111/j.1556-4029.2006.00367.x

49. Buckberry JL (2015) The (mis)use of adult age estimates in osteology. Ann Hum Biol 42:323–331. https://doi.org/10.3109/03014460.2015.1046926

50. Bocquet-Appel JP, Masset C (1982) Farewell to paleodemography. J Hum Evol 11:321–333. https://doi.org/10.1016/S0047-2484(82)80023-7

51. Buk Z, Kordik P, Bruzek J, et al (2012) The age at death assessment in a multi-

ethnic sample of pelvic bones using nature-inspired data mining methods. Forensic Sci Int 220:294.e1-294.e9. https://doi.org/10.1016/j.forsciint.2012.02.019

52. Kotěrová A, Navega D, Štepanovský M, et al (2018) Age estimation of adult human remains from hip bones using advanced methods. Forensic Sci Int 287:163–175. https://doi.org/10.1016/j.forsciint.2018.03.047
53. Navega D, Costa E, Cunha E (2017) Lost in the woods: The value of tree ensemble modelling for adult age-at-death estimation from skeletal degeneration. La Rev Médecine Légale. https://doi.org/10.1016/j.medleg.2017.10.006
54. Konigsberg LW (2015) Multivariate cumulative probit for age estimation using ordinal categorical data. Ann Hum Biol 42:366–376. https://doi.org/10.3109/03014460.2015.1045430
55. Samworth R, Gowland R (2007) Estimation of adult skeletal age-at-death: Statistical assumptions and applications. Int J Osteoarchaeol 17:174–188. https://doi.org/10.1002/oa.867
56. Lucy D, Aykroyd RG, Pollard AM (2002) Nonparametric calibration for age estimation. J R Stat Soc Ser C Appl Stat 51:183–196. https://doi.org/10.1111/1467-9876.00262
57. Konigsberg LW, Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am J Phys Anthropol 89:235–256. https://doi.org/10.1002/ajpa.1330890208
58. Boldsen JL, Milner GR, Konigsberg LW, Wood JW (2002) Transition analysis: a new method for estimating age from the skeletons. In: Hoppa RD, Vaupel JW (eds) Paleodemography: Age distributions from skeletal samples, 1st Ed. Cambridge University Press, pp 73–106
59. Milner GR, Boldsen JL (2012) Transition analysis: A validation study with known-age modern American skeletons. Am J Phys Anthropol 148:98–110. https://doi.org/10.1002/ajpa.22047
60. Steadman DW, Adams BJ, Konigsberg LW (2006) Statistical basis for positive identification in forensic anthropology. Am J Phys Anthropol 131:15–26.

https://doi.org/10.1002/ajpa.20393

61. Corsini MM, Schmitt A, Bruzek J (2005) Aging process variability on the human skeleton: Artificial network as an appropriate tool for age at death assessment. Forensic Sci Int 148:163–167. https://doi.org/10.1016/j.forsciint.2004.05.008

62. Martins R, Oliveira PE, Schmitt A (2012) Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. Forensic Sci Int 219:287.e1-287.e7. https://doi.org/10.1016/j.forsciint.2011.11.031

63. Baccino E, Ubelaker DH, Hayek L-AC, Zerilli A (1999) Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. J Forensic Sci 44:12019J. https://doi.org/10.1520/jfs12019j

64. Brooks S, Suchey JM (1990) Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Hum Evol 5:227–238. https://doi.org/10.1007/BF02437238

65. Hanihara K, Suzuki T (1978) Estimation of age from the pubic symphysis by means of multiple regression analysis. Am J Phys Anthropol 48:233–239. https://doi.org/10.1002/ajpa.1330480218

66. Gilbert BM, McKern TW (1973) A method for aging the female Os pubis. Am J Phys Anthropol 38:31–38. https://doi.org/10.1002/ajpa.1330380109

67. McKern TW, Stewart TD (1957) Skeletal age changes in young American males, analyzed from the standpoint of age identification. Natick, Massachusetts

68. Todd TW (1920) Age changes in the pubic bone. I. The male white pubis. Am J Phys Anthropol 3:285–334. https://doi.org/10.1002/ajpa.1330030301

69. Todd TW (1921) Age changes in the pubic bone. Am J Phys Anthropol 4:1–70. https://doi.org/10.1002/ajpa.1330040102

70. Katz D, Suchey JM (1986) Age determination of the male Os pubis. Am J Phys Anthropol 69:427–435. https://doi.org/10.1002/ajpa.1330690402

71. Stoyanova D, Algee-Hewitt BFB, Slice DE (2015) 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 158:431–440.

https://doi.org/10.1002/ajpa.22797

72. Slice DE, Algee-Hewitt BFB (2015) Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. J Forensic Sci 60:835–843. https://doi.org/10.1111/1556-4029.12778

73. Milner GR, Boldsen JL (2015) Skeletal Age Estimation: Where We Are and Where We Should Go. In: Dirkmaat DC (ed) A Companion to Forensic Anthropology, 1st Ed. Blackwell Publishing Ltd, pp 224–238

74. Milner GR, Boldsen JL (2012) Estimating Age and Sex from the Skeleton, a Paleopathological Perspective. In: Grauer AL (ed) A Companion to Paleopathology, First Edit. Blackwell Publishing Ltd, pp 268–284

75. Vilas-Boas D, Wasterlain SNN, d'Oliveira Coelho J, et al (2019) SPINNE: An app for human vertebral height estimation based on artificial neural networks. Forensic Sci Int 298:121–130. https://doi.org/10.1016/j.forsciint.2019.02.056

76. Scott GR, Pilloud M, Navega D, et al (2018) rASUDAS: A New Web-Based Application for Estimating Ancestry from Tooth Morphology. Forensic Anthropol 1:18–31. https://doi.org/10.5744/fa.2018.0003

77. Navega D, Coelho C, Vicente R, et al (2015) AncesTrees: ancestry estimation with randomized decision trees. Int J Legal Med 129:1145–1153. https://doi.org/10.1007/s00414-014-1050-9

78. Damas S, Wilkinson C, Kahana T, et al (2015) Study on the performance of different craniofacial superimposition approaches (II): Best practices proposal. Forensic Sci Int 257:504–508. https://doi.org/10.1016/j.forsciint.2015.07.045

79. Mesejo P, Martos R, Ibáñez Ó, et al (2020) A Survey on artificial intelligence techniques for biomedical image analysis in skeleton-based forensic human identification. Appl Sci 10:. https://doi.org/10.3390/app10144703

80. Navega D, Costa E, Cunha E (2022) Adult Skeletal Age-At-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis. Biology (Basel) 11:. https://doi.org/10.3390/biology11040532

81. Navega D, Cunha E (2020) Extreme learning machine neural networks for adult

skeletal age-at-death estimation. In: Obertová Z, Stewart A, Cattaneo C (eds) Statistics and Probability in Forensic Anthropology. Elsevier, pp 209–225

82. Navega D (2022) DRNNAGE: Deep random neural networks for adult skeletal age-at-death estimation. https://github.com/dsnavega/DRNNAGE
83. Todd TW (1921) Age changes in the pubic bone. V. Mammalian public metamorphosis. Am J Phys Anthropol 4:333–406. https://doi.org/10.1002/AJPA.1330040402
84. Todd TW, Lyon DW (1924) Cranial Suture Closure, Its Progress and Age Relationship. Part I. Am J Phys Anthropol VII:325–384
85. Todd TW, Lyon DW (1925) Cranial suture closure Part II. Am J Phys Anthropol 9:23–44
86. Singer R (1953) Estimation of age from cranial suture closure: a report on its unreliability. J Forensic Med 1:52–59
87. Brooks S (1955) Skeletal age at death: The reliability of cranial and pubic age indicators. Am J Phys Anthropol 13:567–597. https://doi.org/10.1002/ajpa.1330130403
88. Suchey JM (1979) Problems in the aging of females using theOs pubis. Am J Phys Anthropol 51:467–470. https://doi.org/10.1002/ajpa.1330510319
89. Meindl RS, Lovejoy CO, Mensforth RP, Walker RA (1985) A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. Am J Phys Anthropol 68:29–45. https://doi.org/10.1002/ajpa.1330680104
90. Meindl RS, Lovejoy O (1985) Ectocranial suture closure: A revised method for the determination of skeletal age at death and blind tests of its accuracy. Am J Phys Anthropol 66:57–66
91. İşcan MY, Loth SR (1984) Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. Am J Phys Anthropol 65:147–156
92. İşcan MY, Loth SR, Wright RK (1984) Age estimation from the rib by phase analysis: white males. J Forensic Sci 29:1094–1104.

https://doi.org/10.1520/JFS11776J

93. İşcan MY, Loth SR, Wright RK (1985) Age estimation from the rib by phase analysis: white females. J Forensic Sci 30:853–863. https://doi.org/10.1520/JFS11776J

94. İşcan MY, Loth SR (1986) Determination of age from the sternal rib in white males: a test of the phase method. J Forensic Sci 31:122–132

95. İşcan MY, Loth SR (1986) Determination of age from the sternal rib in white females: a test of the phase method. J Forensic Sci 31:990–999

96. İşcan MY, Loth SR, Wright RK (1987) Racial variation in the sternal extremity of the rib and its effect on age determination. J Forensic Sci 32:452–466. https://doi.org/10.1520/jfs11147j

97. Lovejoy CO, Meindl RS, Pryzbeck TR, Mensforth RP (1985) 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 68:15–28. https://doi.org/10.1002/ajpa.1330680103

98. Kunos CA, Simpson SW, Russell KF, Hershkovitz I (1999) First rib metamorphosis: Its possible utility for human age-at-death estimation. Am J Phys Anthropol 110:303–323. https://doi.org/10.1002/(SICI)1096-8644(199911)110:3<303::AID-AJPA4>3.0.CO;2-O

99. Digangi EA, Bethard JD, Kimmerle EH, Konigsberg LW (2009) A new method for estimating age-at-death from the first rib. Am J Phys Anthropol 138:164–176. https://doi.org/10.1002/ajpa.20916

100. Murray KA, Murray T (1991) A Test of the Auricular Surface Aging Technique. J Forensic Sci 36:13131J. https://doi.org/10.1520/jfs13131j

101. Osborne DL, Simmons TL, Nawrocki SP (2005) Reconsidering the Auricular Surface as an Indicator of Age at Death. J Forensic Sci 49:1–7. https://doi.org/10.1520/jfs2003348

102. Buckberry JL, Chamberlain AT (2002) Age estimation from the auricular surface of the ilium: A revised method. Am J Phys Anthropol 119:231–239.

https://doi.org/10.1002/ajpa.10130

103. Igarashi Y, Uesu K, Wakebe T, Kanazawa E (2005) New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. Am J Phys Anthropol 128:324–339. https://doi.org/10.1002/ajpa.20081

104. Falys CG, Schutkowski H, Weston DA (2006) Auricular surface aging: Worse than expected? A test of the revised method on a documented historic skeletal assemblage. Am J Phys Anthropol 130:508–513. https://doi.org/10.1002/ajpa.20382

105. Mulhern DM, Jones EB (2005) Test of revised method of age estimation from the auricular surface of the ilium. Am J Phys Anthropol 126:61–65. https://doi.org/10.1002/ajpa.10410

106. Moraitis K, Zorba E, Eliopoulos C, Fox SC (2014) A Test of the Revised Auricular Surface Aging Method on a Modern European Population. J Forensic Sci 59:188–194. https://doi.org/10.1111/1556-4029.12303

107. Hens SM, Belcastro MG (2012) Auricular surface aging: A blind test of the revised method on historic Italians from Sardinia. Forensic Sci Int 214:209.e1-209.e5. https://doi.org/10.1016/j.forsciint.2011.07.043

108. Hens SM, Rastelli E, Belcastro G (2008) Age estimation from the human os coxa: A test on a documented Italian collection. J Forensic Sci 53:1040–1043. https://doi.org/10.1111/j.1556-4029.2008.00818.x

109. Lottering N, MacGregor DM, Meredith M, et al (2013) Evaluation of the suchey-brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. Am J Phys Anthropol 150:386–399. https://doi.org/10.1002/ajpa.22213

110. Villa C, Buckberry J, Cattaneo C, et al (2015) Quantitative analysis of the morphological changes of the pubic symphyseal face and the auricular surface and implications for age at death estimation. J Forensic Sci 60:556–565. https://doi.org/10.1111/1556-4029.12689

111. Kotěrová A, Králík V, Rmoutilová R, et al (2019) Impact of 3D surface scanning protocols on the Os coxae digital data: Implications for sex and age-at-death

assessment. J Forensic Leg Med 68:. https://doi.org/10.1016/j.jflm.2019.101866

112. Stoyanova DK, Algee-Hewitt BFB, Kim J, Slice DE (2017) A Computational Framework for Age-at-Death Estimation from the Skeleton: Surface and Outline Analysis of 3D Laser Scans of the Adult Pubic Symphysis. J Forensic Sci 62:1434–1444. https://doi.org/10.1111/1556-4029.13439

113. Kim J, Algee-Hewitt BFB, Stoyanova DK, et al (2019) Testing Reliability of the Computational Age-At-Death Estimation Methods between Five Observers Using Three-Dimensional Image Data of the Pubic Symphysis,. J Forensic Sci 64:507–518. https://doi.org/10.1111/1556-4029.13842

114. Fojas CL, Kim J, Minsky-Rowland JD, Algee-Hewitt BFB (2018) Testing inter-observer reliability of the Transition Analysis aging method on the William M. Bass forensic skeletal collection. Am J Phys Anthropol 165:183–193. https://doi.org/10.1002/ajpa.23342

115. Berg GE (2008) Pubic bone age estimation in adult women. J Forensic Sci 53:569–577. https://doi.org/10.1111/j.1556-4029.2008.00712.x

116. Hartnett KM (2010) Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample-Part II: Sternal end of the fourth rib. J Forensic Sci 55:1145–1151. https://doi.org/10.1111/j.1556-4029.2010.01399.x

117. Chen X, Zhang Z, Tao L (2008) Determination of male age at death in Chinese Han population: Using quantitative variables statistical analysis from pubic bones. Forensic Sci Int 175:36–43. https://doi.org/10.1016/j.forsciint.2007.04.231

118. Chen XP, Zhang ZY, Zhu GY, Tao LY (2011) Determining the age at death of females in the Chinese Han population: Using quantitative variables and statistical analysis from pubic bones. Forensic Sci Int 210:278.e1-278.e8. https://doi.org/10.1016/j.forsciint.2011.04.024

119. Fleischman JM (2013) A comparative assessment of the chen et al. and suchey-brooks pubic aging methods on a North American sample. J Forensic Sci 58:311–323. https://doi.org/10.1111/1556-4029.12061

120. Castillo A, Galtés I, Crespo S, Jordana X (2020) Technical note: preliminary insight

into a new method for age-at-death estimation from the pubic symphysis. Int J Legal Med. https://doi.org/10.1007/s00414-020-02434-6

121. Passalacqua N V. (2009) Forensic age-at-death estimation from the human sacrum. J Forensic Sci 54:255–262. https://doi.org/10.1111/j.1556-4029.2008.00977.x

122. Colarusso T (2016) A Test of the Passalacqua Age at Death Estimation Method Using the Sacrum. J Forensic Sci 61:22–29. https://doi.org/10.1111/1556-4029.12967

123. Falys CG, Prangle D (2015) Estimating age of mature adults from the degeneration of the sternal end of the clavicle. Am J Phys Anthropol 156:203–214. https://doi.org/10.1002/ajpa.22639

124. Botha D, Pretorius S, Myburgh J, Steyn M (2016) Age estimation from the acetabulum in South African black males. Int J Legal Med 130:809–817. https://doi.org/10.1007/s00414-015-1299-7

125. Winburn AP (2019) Validation of the Acetabulum As a Skeletal Indicator of Age at Death in Modern European-Americans. J Forensic Sci 64:989–1003. https://doi.org/10.1111/1556-4029.13972

126. Rougé-Maillart C, Vielle B, Jousset N, et al (2009) 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 188:91–95. https://doi.org/10.1016/j.forsciint.2009.03.019

127. Ferrant O, Rougé-Maillart C, Guittet L, et al (2009) Age at death estimation of adult males using coxal bone and CT scan: A preliminary study. Forensic Sci Int 186:14–21. https://doi.org/10.1016/j.forsciint.2008.12.024

128. San-Millán M, Rissech C, Turbón D (2017) New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. Int J Legal Med 131:501–525. https://doi.org/10.1007/s00414-016-1406-4

129. San-Millán M, Rissech C, Turbón D (2019) Application of the recent SanMillán–Rissech acetabular adult aging method in a North American sample. Int J Legal

Med 133:909–920. https://doi.org/10.1007/s00414-019-02005-4

130. Rissech C, Estabrook GF, Cunha E, Malgosa A (2006) Using the acetabulum to estimate age at death of adult males. J Forensic Sci 51:213–229. https://doi.org/10.1111/j.1556-4029.2006.00060.x

131. Rissech C, Estabrook GF, Cunha E, Malgosa A (2007) Estimation of age-at-death for adult males using the acetabulum, applied to four Western European populations. J Forensic Sci 52:774–778. https://doi.org/10.1111/j.1556-4029.2007.00486.x

132. Rougé-Maillart C, Telmon N, Rissech C, et al (2004) The Determination of Male Adult Age at Death by Central and Posterior Coxal Analysis A Preliminary Study. J Forensic Sci 49:1–7. https://doi.org/10.1520/jfs2002056

133. Rissech C, Wilson J, Winburn AP, et al (2012) A comparison of three established age estimation methods on an adult Spanish sample. Int J Legal Med 126:145–155. https://doi.org/10.1007/s00414-011-0586-1

134. Mays S (2014) A Test of a Recently Devised Method of Estimating Skeletal Age at Death using Features of the Adult Acetabulum. J Forensic Sci 59:184–187. https://doi.org/10.1111/1556-4029.12293

135. Mays S (2012) An investigation of age-related changes at the acetabulum in 18th-19th century ad adult skeletons from Christ Church Spitalfields, London. Am J Phys Anthropol 149:485–492. https://doi.org/10.1002/ajpa.22146

136. Navega D, Godinho M, Cunha E, Ferreira MT (2018) A test and analysis of Calce (2012) method for skeletal age-at-death estimation using the acetabulum in a modern skeletal sample. Int J Legal Med 132:1447–1455. https://doi.org/10.1007/s00414-018-1902-9

137. Calce SE, Rogers TL (2011) Evaluation of age estimation technique: Testing traits of the acetabulum to estimate age at death in adult males. J Forensic Sci 56:302–311. https://doi.org/10.1111/j.1556-4029.2011.01700.x

138. Calce SE (2012) A new method to estimate adult age-at-death using the acetabulum. Am J Phys Anthropol 148:11–23. https://doi.org/10.1002/ajpa.22026

139. Rougé-Maillart C, Jousset N, Vielle B, et al (2007) Contribution of the study of acetabulum for the estimation of adult subjects. Forensic Sci Int 171:103–110. https://doi.org/10.1016/j.forsciint.2006.10.007

140. Rivera-Sandoval J, Monsalve T, Cattaneo C (2018) A test of four innominate bone age assessment methods in a modern skeletal collection from Medellin, Colombia. Forensic Sci Int 282:. https://doi.org/10.1016/j.forsciint.2017.11.003

141. Listi GA, Manhein MH (2012) The Use of Vertebral Osteoarthritis and Osteophytosis in Age Estimation. J Forensic Sci 57:1537–1540. https://doi.org/10.1111/j.1556-4029.2012.02152.x

142. Listi GA (2016) The Use of Entheseal Changes in the Femur and Os Coxa for Age Assessment. J Forensic Sci 61:12–18. https://doi.org/10.1111/1556-4029.12905

143. Alves-Cardoso F, Assis S (2018) Can osteophytes be used as age at death estimators? Testing correlations in skeletonized human remains with known age-at-death. Forensic Sci Int 288:59–66. https://doi.org/10.1016/j.forsciint.2018.04.034

144. Winburn AP, Stock MK (2019) Reconsidering osteoarthritis as a skeletal indicator of age at death. Am J Phys Anthropol 170:459–473. https://doi.org/10.1002/ajpa.23914

145. Milella M, Belcastro MG, Mariotti V, Nikita E (2020) Estimation of adult age-at-death from entheseal robusticity: A test using an identified Italian skeletal collection. Am J Phys Anthropol 173:190–199. https://doi.org/10.1002/ajpa.24083

146. Santos A (1995) Certezas e incertezas sobre a idade à morte. Universidade de Coimbra

147. Schmitt A, Murail P, Cunha E, Rougé D (2002) Variability of the Pattern of Aging on the Human Skeleton: Evidence from Bone Indicators and Implications on Age at Death Estimation. J Forensic Sci 47:15551J. https://doi.org/10.1520/jfs15551j

148. Komar DA, Grivas C (2008) Manufactured populations: What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. Am J Phys Anthropol 137:224–233. https://doi.org/10.1002/ajpa.20858

149. Aykroyd RG, Lucy D, Pollard AM, Solheim T (1997) Technical note: Regression analysis in adult age estimation. Am J Phys Anthropol 104:259–265. https://doi.org/10.1002/(SICI)1096-8644(199710)104:2<259::AID-AJPA11>3.0.CO;2-Z

150. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH (2008) Estimation and evidence in forensic anthropology: Age-at-death. J Forensic Sci 53:541–557. https://doi.org/10.1111/j.1556-4029.2008.00710.x

151. Stull KE, Chu EY, Corron LK, Price MH (2022) Subadult Age Estimation Using the Mixed Cumulative Probit and a Contemporary United States Population. Forensic Sci 2:741–779. https://doi.org/10.3390/forensicsci2040055

152. Ferembach D, Schwindezky I, Stoukal M (1980) Recommendations for age and sex diagnoses of skeletons. J Hum Evol 9:517–549. https://doi.org/10.1016/0047-2484(80)90061-5

153. Lovejoy CO, Meindl RS, Mensforth RP, Barton TJ (1985) Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. Am J Phys Anthropol 68:1–14. https://doi.org/10.1002/ajpa.1330680102

154. Jooste N, L'Abbé EN, Pretorius S, Steyn M (2016) Validation of transition analysis as a method of adult age estimation in a modern South African sample. Forensic Sci Int 266:580.e1-580.e7. https://doi.org/10.1016/j.forsciint.2016.05.020

155. Milner GR, Boldsen JL, Ousley SD, et al (2021) Great expectations: The rise, fall, and resurrection of adult skeletal age estimation. In: Algee-Hewitt BFB, Kim J (eds) Remodeling Forensic Skeletal Age. Elsevier, pp 139–154

156. Anderson DT, Havens TC, Wagner C, et al (2012) Sugeno fuzzy integral generalizations for Sub-normal Fuzzy set-valued inputs. IEEE Int Conf Fuzzy Syst 10–15. https://doi.org/10.1109/FUZZ-IEEE.2012.6250827

157. Anderson DT, Keller JM, Anderson M, Wescott DJ (2011) Linguistic description of adult skeletal age-at-death estimations from fuzzy integral acquired fuzzy sets. IEEE Int Conf Fuzzy Syst 2274–2281. https://doi.org/10.1109/FUZZY.2011.6007421

158. Anderson MF, Anderson DT, Wescott DJ (2010) Estimation of adult skeletal age-

at-death using the sugeno fuzzy integral. Am J Phys Anthropol 142:30–41. https://doi.org/10.1002/ajpa.21190

159. Baccino E, Sinfield L, Colomb S, et al (2014) Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. Forensic Sci Int 244:247–251. https://doi.org/10.1016/j.forsciint.2014.09.005
160. Kemkes-Grottenthaler A (2002) Aging through the ages: historical perspectives on age indicators methods. In: Hoppa RD, Vaupel JW (eds) Paleodemography: Age distributions from skeletal samples, First Edit. Cambridge University Press, pp 48–72
161. Shugen W (2002) Framework of pattern recognition model based on the cognitive psychology. Geo-Spatial Inf Sci 5:74–78. https://doi.org/10.1007/BF02833890
162. Solso RL, Maclin OH, MacLin MK (2014) Cognitive Psychology: Pearson New International Edition. Pearson Education Limited, Essex
163. Eysenck MW, Keane MT (2015) Cognitive Psychology: A Student's Handbook, 7th ed. Psychology Press, Taylor and Francis Group, New York
164. Neisser U, Niesser U (2014) Cognitive Psychology: Classic Edition. Psychology Press, Taylor and Francis Group, New York
165. Friedenberg J, Silverman G (2006) Cognitive Sciences: An Introduction to the Study of Mind. Sage Publications Inc., London
166. Sternberg RJ, Sternberg K (2012) Cognitive Psychology, 6th Editio. Wadsworth, Cengage Learning, Belmont
167. Shirley NR, Ramirez Montes PA (2015) Age estimation in forensic anthropology: Quantification of observer error in phase versus component-based methods. J Forensic Sci 60:107–111. https://doi.org/10.1111/1556-4029.12617
168. Perizonius WRK (1984) Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (a.d. 1883-1909). J Hum Evol 13:201–216. https://doi.org/10.1016/S0047-2484(84)80065-2
169. Mann RW, Jantz RL, Bass WM, Willey PS (1991) Maxillary suture obliteration: a visual method for estimating skeletal age. J Forensic Sci 36:781–91.

https://doi.org/10.1520/jfs13088j

170. Mann RW, Symes SA, Bass WM (1987) Maxillary Suture Obliteration: Aging the Human Skeleton Based on Intact or Fragmentary Maxilla. J Forensic Sci 32:12337J

171. Acsadi J, Nemeskeri G (1970) History of Human Life Span and Mortality. Académiai Kiadó, Budapest

172. Masset C (1989) Age estimation on the basis of cranial sutures. In: İşcan MY (ed) Age Markers in the Human Skeleton. Charles C Thomas, Springfield, pp 71–103

173. Ríos L, Weisensee K, Rissech C (2008) Sacral fusion as an aid in age estimation. Forensic Sci Int 180:111.e1-111.e7. https://doi.org/10.1016/j.forsciint.2008.06.010

174. Belcastro MG, Rastelli E, Mariotti V (2008) Variation of the degree of sacral vertebral body fusion in adulthood in two European modern skeletal collections. Am J Phys Anthropol 135:149–160. https://doi.org/10.1002/ajpa.20716

175. Snodgrass JJ (2004) Sex Differences and Aging of the Vertebral Column. J Forensic Sci 49:1–6. https://doi.org/10.1520/jfs2003198

176. Watanabe S, Terazawa K (2006) Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. Leg Med 8:156–160. https://doi.org/10.1016/j.legalmed.2006.01.001

177. Albert M, Mulhern D, Torpey MA, Boone E (2010) Age estimation using thoracic and first two lumbar vertebral ring epiphyseal union. J Forensic Sci 55:287–294. https://doi.org/10.1111/j.1556-4029.2009.01307.x

178. Milella M, Giovanna Belcastro M, Zollikofer CPE, Mariotti V (2012) The effect of age, sex, and physical activity on entheseal morphology in a contemporary Italian skeletal collection. Am J Phys Anthropol 148:379–388. https://doi.org/10.1002/ajpa.22060

179. Calce SE, Kurki HK, Weston DA, Gould L (2018) Effects of osteoarthritis on age-at-death estimates from the human pelvis. Am J Phys Anthropol 167:. https://doi.org/10.1002/ajpa.23595

180. Calce SE, Kurki HK, Weston DA, Gould L (2017) Principal component analysis in the evaluation of osteoarthritis. Am J Phys Anthropol 162:476–490.

https://doi.org/10.1002/ajpa.23130

181. Brennaman AL, Love KR, Bethard JD, Pokines JT (2017) A Bayesian Approach to Age-at-Death Estimation from Osteoarthritis of the Shoulder in Modern North Americans. J Forensic Sci 62:573–584. https://doi.org/10.1111/1556-4029.13327

182. Calce SE, Kurki HK, Weston DA, Gould L (2018) The relationship of age, activity, and body size on osteoarthritis in weight-bearing skeletal regions. Int J Paleopathol 22:45–53. https://doi.org/10.1016/j.ijpp.2018.04.001

183. Buikstra JE, Ubelaker DH (1994) Standards for Data Collection from Human Skeletal Remains. Arkansas Archeological Survey Research Series No. 44, Fayetteville

184. Henderson CY, Mariotti V, Pany-Kucera D, et al (2013) Recording Specific Entheseal Changes of Fibrocartilaginous Entheses: Initial Tests Using the Coimbra Method. Int J Osteoarchaeol 23:152–162. https://doi.org/10.1002/oa.2287

185. Langley-Shirley N, Jantz RL (2010) A bayesian approach to age estimation in modern Americans from the clavicle. J Forensic Sci 55:571–583. https://doi.org/10.1111/j.1556-4029.2010.01089.x

186. Owings WPA, Myers SJ, Webb PAO, Suchey JM (2005) Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. Am J Phys Anthropol 68:457–466. https://doi.org/10.1002/ajpa.1330680402

187. Cardoso HFV V (2008) 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

188. Sashin D (1930) A critical analysis pf the anatomy and the pathologic changes of the sacro-iliac joints. J Bone Jt Surg 12:891–910. https://doi.org/10.1097@00401515

189. Schunke GB (1938) The anatomy and development of the sacro-iliac joint in man. Anat Rec 72:313–331. https://doi.org/10.1002/ar.1090720306

190. Stull KE, James DM (2010) Determination of age at death using the acetabulum of the os coxa. In: Latham K, Finnegan M (eds) Age Estimation of the Human

Skeleton, 1st ed. Charles C. Thomas, pp 134–146

191. Cohen J (1960) A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20:37–46. https://doi.org/10.1177/001316446002000104
192. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33:613–619. https://doi.org/10.1177/001316447303300309
193. Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 70:213–220. https://doi.org/10.1037/h0026256
194. Kendall MG, Smith BB (1939) The Problem of m Rankings. Ann Math Stat 10:275–287. https://doi.org/10.1214/aoms/1177732186
195. Steadman DW (2018) Who needs data? I've got experience! Hum Biol 90:77–82. https://doi.org/10.13110/humanbiology.90.1.05
196. Algee-Hewitt BFB, Kim J, Hughes CE, et al (2018) Thinking computationally about forensics: Anthropological perspectives on advancements in technologies, data, and algorithms. Hum Biol 90:5–10. https://doi.org/10.13110/humanbiology.90.1.04
197. Usher BM (2002) Reference samples: the first step in linking biology and age in the human skeleton. In: Hoppa RD, Vaupel JW (eds) Paleodemography. Cambridge University Press, pp 29–47
198. Adalian P (2020) General considerations about data and selection of statistical approaches. In: Obertová Z, Alistair S, Cattaneo C (eds) Statistics and Probability in Forensic Anthropology. Elsevier, pp 59–72
199. Madrigal L (2012) Statistics for Anthropology, 2nd ed. Cambridge University Press
200. Cunha E, Wasterlain S (2008) The Coimbra identified osteological collections. In: Grupe G, Peters J (eds) Skeletal series and their socio-economic context (Documenta Archaeobiologiae; Bd. 5), 1st ed. Verlag Marie Leidorf GmbH, Rahden/Westf., pp 23–34
201. Ferreira MT, Vicente R, Navega D, et al (2014) A new forensic collection housed at the University of Coimbra, Portugal: The 21st century identified skeletal collection.

Forensic Sci Int 245:202.e1-202.e5. https://doi.org/10.1016/j.forsciint.2014.09.021

202. Ferreira MT, Coelho C, Makhoul C, et al (2021) New data about the 21st Century Identified Skeletal Collection (University of Coimbra, Portugal). Int J Legal Med 135:1087–1094. https://doi.org/10.1007/s00414-020-02399-6
203. Pokines JT, Symes SA (2014) Manual of Forensic Taphonomy, 1st ed. CRC Press, Boca Raton
204. Nikita E, Nikitas P (2019) Skeletal age-at-death estimation: Bayesian versus regression methods. Forensic Sci Int 297:56–64. https://doi.org/10.1016/j.forsciint.2019.01.033
205. Langley NR, Meadows Jantz L, Ousley SD, et al (2018) Data Collection Procedures for Forensic Skeletal Material 2.0, 2nd ed
206. DiGangi E, Moore M (2013) Research Methods in Human Skeletal Biology. Elsevier
207. Kenyhercz M, Passalacqua N, Hefner J (2019) Missing Data Imputation Using Morphoscopic Traits and Their Performance in the Estimation of Ancestry. Forensic Anthropol 2:10–12. https://doi.org/10.5744/fa.2019.1015
208. Howells WW (1973) Cranial variation in man: a study by multivariate analysis of patterns of difference among recent human populations. Peabody Museum of Archaeology and Ethnology, Harvard University
209. Beretta L, Santaniello A (2016) Nearest neighbor imputation algorithms: a critical evaluation. BMC Med Inform Decis Mak 16:74. https://doi.org/10.1186/s12911-016-0318-z
210. Stekhoven DJ, Bühlmann P (2012) Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics 28:112–118. https://doi.org/10.1093/bioinformatics/btr597
211. Tang F, Ishwaran H (2017) Random Forest Missing Data Algorithms
212. Liao SG, Lin Y, Kang DD, et al (2014) Missing value imputation in high-dimensional phenomic data: Imputable or not, and how? BMC Bioinformatics 15:1–12. https://doi.org/10.1186/s12859-014-0346-6
213. Cramér H (1999) Mathematical Methods of Statistics. Princeton university press

214. Kolmogorov A (1933) Sulla determinazione empirica di una lgge di distribuzione. Inst Ital Attuari, Giorn 4:83–91

215. Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions. Ann Math Stat 19:279–281

216. Bhapkar VP (1966) A Note on the Equivalence of Two Test Criteria for Hypotheses in Categorical Data. J Am Stat Assoc 61:228–235. https://doi.org/10.1080/01621459.1966.10502021

217. Fisher RA (1925) Statistical Methods for Research Workers. Oliver and Boyd

218. Fieuws S, Willems G, Larsen-Tangmose S, et al (2016) Obtaining appropriate interval estimates for age when multiple indicators are used: evaluation of an ad-hoc procedure. Int J Legal Med 130:489–499. https://doi.org/10.1007/s00414-015-1200-8

219. Lynch J, Stephan C (2018) Computational Tools in Forensic Anthropology: The Value of Open-Source Licensing as a Standard. Forensic Anthropol 1:228–243. https://doi.org/10.5744/fa.2018.0025

220. European Organization For Nuclear Research, OpenAIRE (2013) Zenodo. https://www.zenodo.org/

221. Stull KE, L'Abbé EN, Ousley SD (2014) Using multivariate adaptive regression splines to estimate subadult age from diaphyseal dimensions. Am J Phys Anthropol 154:376–386. https://doi.org/10.1002/ajpa.22522

222. Watt J, Borhani R, Katsaggelos AK (2020) Machine learning refined: Foundations, algorithms, and applications. Cambridge University Press

223. Flach P (2012) Machine learning: the art and science of algorithms that make sense of data. Cambridge university press

224. Theodoridis S (2015) Machine learning: a Bayesian and optimization perspective. Academic press

225. Jung A (2022) Machine Learning: The Basics. Springer Nature

226. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, Cambridge, MA

227. Breiman L (2001) Statistical modeling: The two cultures. Stat Sci 16:199–215. https://doi.org/10.1214/ss/1009213726

228. Stanley KO, Miikkulainen R (2002) Evolving neural networks through augmenting topologies. Evol Comput 10:99–127. https://doi.org/10.1162/106365602320169811

229. Assunção F, Lourenço N, Ribeiro B, Machado P (2021) Fast-DENSER: Fast Deep Evolutionary Network Structured Representation. SoftwareX 14:100694. https://doi.org/10.1016/J.SOFTX.2021.100694

230. Assunção F, Lourenço N, Machado P, Ribeiro B (2019) Fast DENSER: Efficient Deep NeuroEvolution. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11451 LNCS:197–212. https://doi.org/10.1007/978-3-030-16670-0_13

231. Assunção F, Lourenço N, Machado P, Ribeiro B (2019) DENSER: deep evolutionary network structured representation. Genet Program Evolvable Mach 20:5–35. https://doi.org/10.1007/S10710-018-9339-Y

232. Assunção F, Lourenço N, Machado P, Ribeiro B (2019) Fast-DENSER++: Evolving Fully-Trained Deep Artificial Neural Networks

233. Assunção F, Lourenço N, Ribeiro B, Machado P (2020) Incremental Evolution and Development of Deep Artificial Neural Networks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 12101 LNCS:35–51. https://doi.org/10.1007/978-3-030-44094-7_3

234. Scardapane S, Wang D (2017) Randomness in neural networks: an overview. Wiley Interdiscip Rev Data Min Knowl Discov 7:. https://doi.org/10.1002/widm.1200

235. Gallicchio C, Scardapane S (2020) Deep Randomized Neural Networks. Stud Comput Intell 896:43–68. https://doi.org/10.1007/978-3-030-43883-8_3

236. Schmidt WF, Kraaijveld MA, Duin RPW (1992) Feedforward neural networks with random weights. In: Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems. IEEE Comput. Soc. Press, pp 1–4

237. Pao YH, Takefuji Y (1992) Functional-Link Net Computing: Theory, System

Architecture, and Functionalities. Computer (Long Beach Calif) 25:76–79. https://doi.org/10.1109/2.144401

238. Broomhead DS., Lowe D (1988) Multivariable functional interpolation and adaptive networks. Complex Syst 2:321–355

239. Pao YH, Park GH, Sobajic DJ (1994) Learning and generalization characteristics of the random vector functional-link net. Neurocomputing 6:163–180. https://doi.org/10.1016/0925-2312(94)90053-1

240. Igelnik B, Pao YH (1995) Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net. IEEE Trans Neural Networks 6:1320–1329. https://doi.org/10.1109/72.471375

241. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: Theory and applications. Neurocomputing 70:489–501. https://doi.org/10.1016/j.neucom.2005.12.126

242. Huang G, Huang G Bin, Song S, You K (2015) Trends in extreme learning machines: A review. Neural Networks 61:32–48

243. Huang G Bin (2014) An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. Cognit Comput 6:376–390. https://doi.org/10.1007/s12559-014-9255-2

244. Huang G-B (2015) What are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle. Cognit Comput 7:263–278. https://doi.org/10.1007/s12559-015-9333-0

245. Wang LP, Wan CR (2008) Comments on "The extreme learning machine." IEEE Trans Neural Networks 19:1494–1495. https://doi.org/10.1109/TNN.2008.2002273

246. Shao Z, Er MJ (2016) Efficient Leave-One-Out Cross-Validation-based Regularized Extreme Learning Machine. Neurocomputing 194:260–270. https://doi.org/10.1016/j.neucom.2016.02.058

247. Wang D, Wang P, Shi J (2018) A fast and efficient conformal regressor with regularized extreme learning machine. Neurocomputing 304:1–11. https://doi.org/10.1016/j.neucom.2018.04.012

248. Tissera MD, McDonnell MD (2016) Deep extreme learning machines: Supervised autoencoding architecture for classification. Neurocomputing 174:42–49. https://doi.org/10.1016/j.neucom.2015.03.110

249. Tang J, Deng C, Huang G-B (2016) Extreme Learning Machine for Multilayer Perceptron. IEEE Trans Neural Networks Learn Syst 27:809–821. https://doi.org/10.1109/TNNLS.2015.2424995

250. Zhou H, Huang G Bin, Lin Z, et al (2015) Stacked Extreme Learning Machines. IEEE Trans Cybern 45:2013–2025. https://doi.org/10.1109/tcyb.2014.2363492

251. Yu W, Zhuang F, He Q, Shi Z (2015) Learning deep representations via extreme learning machines. Neurocomputing 149:308–315. https://doi.org/10.1016/j.neucom.2014.03.077

252. Shi Q, Katuwal R, Suganthan PN, Tanveer M (2021) Random vector functional link neural network based ensemble deep learning. Pattern Recognit 117:107978. https://doi.org/10.1016/j.patcog.2021.107978

253. McDonnell MD, Tissera MD, Vladusich T, et al (2015) Fast, simple and accurate handwritten digit classification by training shallow neural network classifiers with the "Extreme learning machine" algorithm. PLoS One 10:1–20. https://doi.org/10.1371/journal.pone.0134254

254. Huang G-B, Zhou H, Ding X, Zhang X (2011) Extreme Learning Machine for Regression and Multiclass Classification. IEEE Trans Syst Man, Cybern Part B 42:513–529. https://doi.org/10.1109/tsmcb.2011.2168604

255. Bartlett PL (1998) 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 44:525–536. https://doi.org/10.1109/18.661502

256. Bartlett PL (1996) For valid generalization, the size of the weights is more important than the size of the network. Adv Neural Inf Process Syst 134--140

257. Allen DM (1974) The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics 16:125. https://doi.org/10.2307/1267500

258. Shrestha DL, Solomatine DP (2006) Machine learning approaches for estimation of prediction interval for the model output. Neural Networks 19:225–235. https://doi.org/10.1016/j.neunet.2006.01.012

259. Milborrow S (2021) Variance models in earth

260. Vovk V, Gammerman A, Shafer G (2005) Algorithmic learning in a random world. Springer-Verlag, New York

261. Shafer G, Vovk V (2007) A tutorial on conformal prediction

262. Lei J, G'Sell M, Rinaldo A, et al (2018) Distribution-Free Predictive Inference for Regression. J Am Stat Assoc 113:1094–1111. https://doi.org/10.1080/01621459.2017.1307116

263. Papadopoulos H, Vovk V, Gammerman A (2007) Conformal prediction with neural networks. Proc - Int Conf Tools with Artif Intell ICTAI 2:388–395. https://doi.org/10.1109/ICTAI.2007.47

264. Papadopoulos H, Haralambous H (2011) Reliable prediction intervals with regression neural networks. Neural Networks 24:842–851. https://doi.org/10.1016/j.neunet.2011.05.008

265. Norinder U, Carlsson L, Boyer S, Eklund M (2015) Introducing conformal prediction in predictive modeling for regulatory purposes. A transparent and flexible alternative to applicability domain determination. Regul Toxicol Pharmacol 71:279–284. https://doi.org/10.1016/j.yrtph.2014.12.021

266. Vovk V (2015) Cross-conformal predictors. Ann Math Artif Intell 74:9–28. https://doi.org/10.1007/s10472-013-9368-4

267. Gammerman A, Vovk V (2006) Hedging Predictions in Machine Learning: The Second Computer Journal Lecture. Comput J 50:151–163. https://doi.org/10.1093/comjnl/bxl065

268. Norinder U, Carlsson L, Boyer S, et al (2014) Introducing Conformal Prediction in Predictive Modeling. A Transparent and Flexible Alternative To Applicability Domain Determination. J Chem Inf Model 54:1596–1603. https://doi.org/10.1021/ci5001168

269. Kuhn M, Johnson K (2013) Applied predictive modeling. Springer New York

270. Teixeira F, Cunha E (2021) Aging the elderly: Does the skull tell us something about age at death? In: Algee-Hewitt BFB, Kim J (eds) Remodeling Forensic Skeletal Age. Academic Press, pp 75–97

271. Lourenço M, Cunha E (2020) Can we still use cranial sutures to estimate age at death of individuals after age 50. Braz J Forensic Anthr \& Leg Med 1:5–28

272. Getz SM (2020) The use of transition analysis in skeletal age estimation. WIREs Forensic Sci 2:1–11. https://doi.org/10.1002/wfs2.1378

273. Miranker M (2016) A Comparison of Different Age Estimation Methods of the Adult Pelvis. J Forensic Sci 61:1173–1179. https://doi.org/10.1111/1556-4029.13130

274. Merritt CE (2017) Inaccuracy and bias in adult skeletal age estimation: Assessing the reliability of eight methods on individuals of varying body sizes. Forensic Sci Int 275:315.e1-315.e11. https://doi.org/10.1016/j.forsciint.2017.03.003

275. Hagelthorn CL, Alblas A, Greyling L (2019) The accuracy of the Transition Analysis of aging on a heterogenic South African population. Forensic Sci Int 297:370.e1-370.e5. https://doi.org/10.1016/j.forsciint.2019.02.012

276. Valsecchi A, Irurita Olivares J, Mesejo P (2019) Age estimation in forensic anthropology: methodological considerations about the validation studies of prediction models. Int J Legal Med 1–10. https://doi.org/10.1007/s00414-019-02064-7

277. Jooste N, Pretorius S, Steyn M (2021) Performance of three mathematical models for estimating age-at-death from multiple indicators of the adult skeleton. Int J Legal Med 9–11. https://doi.org/10.1007/s00414-021-02727-4

278. Sanderson C, Curtin R (2016) Armadillo: a template-based C++ library for linear algebra. J Open Source Softw 1:26. https://doi.org/10.21105/joss.00026

279. d'Oliveira Coelho J, Curate F, Navega D (2020) Osteomics: Decision support systems for forensic anthropologists. In: Obertová Z, Stewart A, Cattaneo C (eds) Statistics and Probability in Forensic Anthropology. Elsevier, pp 259–273

280. Git (2022) Free and Open Source Distributed Version Control System. https: //git-

scm.com

281. GitHub (2022) Git Repository Hosting Service. https://github.com/
282. R Core Team (2022) R: A language and environment for statistical computing. https://www.r-project.org/
283. Chang W, Cheng J, Allaire J, Xie Y (2022) shiny: Web Application Framework for R. https://cran.r-project.org/package=shiny
284. Posit (2022) shinyserver: A Web Server for R Web Applications. https://posit.co/products/open-source/shinyserver/
285. Eddelbuettel D, François R (2011) Rcpp: Seamless R and C++ Integration. J Stat Softw 40:1–18. https://doi.org/10.18637/jss.v040.i08
286. Eddelbuettel D (2013) Seamless R and C++ Integration with Rcpp. Springer, New York
287. Eddelbuettel D, Balamuta JJ (2018) Extending R with C++: A Brief Introduction to Rcpp. Am Stat 72:28–36. https://doi.org/10.1080/00031305.2017.1375990
288. Eddelbuettel D, Sanderson C (2014) RcppArmadillo: Accelerating R with high-performance C++ linear algebra. Comput Stat Data Anal 71:1054–1063
289. Sanderson C, Curtin R (2018) A User-Friendly Hybrid Sparse Matrix Class in C++. In: Lecture Notes in Computer Science. pp 422–430
290. Burkart N, Huber MF (2021) A survey on the explainability of supervised machine learning. J Artif Intell Res 70:245–317. https://doi.org/10.1613/JAIR.1.12228
291. Molnar C (2022) Interpretable Machine Learning, 2nd ed
292. Lundberg SM, Allen PG, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions
293. Murdoch WJ, Singh C, Kumbier K, et al (2019) Definitions, methods, and applications in interpretable machine learning. Proc Natl Acad Sci U S A 116:22071–22080. https://doi.org/10.1073/pnas.1900654116
294. Biecek P, Burzykowski T (2021) Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models. Chapman and Hall/CRC, New York
295. Rudin C (2019) Stop explaining black box machine learning models for high stakes

decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x

296. Zuber V, Strimmer K (2011) High-dimensional regression and variable selection using CAR scores. Stat Appl Genet Mol Biol 10:. https://doi.org/10.2202/1544-6115.1730

297. Kessy A, Lewin A, Strimmer K (2018) Optimal Whitening and Decorrelation. Am Stat 72:309–314. https://doi.org/10.1080/00031305.2016.1277159

[*This page has been intentionally left blank*]

Appendix A

Sex-related inferential analysis

Table A.1 Cramér's V for assessment skeletal morphology and sex association on bilateral traits.

Trait	Side	Statistic	p-value	df	n
SC01	Left	0.000	0.563	1	483
	Right	0.000	0.472	1	488
HM01	Left	0.033	0.216	1	479
	Right	0.000	0.853	1	474
HM02	Left	0.000	0.491	1	475
	Right	0.034	0.215	1	469
HM03	Left	0.043	0.174	1	464
	Right	0.000	0.667	1	454
HM04	Left	0.000	0.600	1	476
	Right	0.049	0.147	1	469
HM05	Left	0.000	0.406	1	438
	Right	0.000	0.681	1	433
HM06	Left	0.056	0.122	1	440
	Right	0.000	0.320	1	431
UL01	Left	0.069	0.071	1	475
	Right	0.100	0.016	1	476
UL02	Left	0.031	0.231	1	449
	Right	0.056	0.120	1	443
RD01	Left	0.000	0.685	1	466
	Right	0.000	0.616	1	465
RD02	Left	0.000	0.981	1	476
	Right	0.000	0.700	1	468
OC01	Left	0.000	0.485	1	433
	Right	0.000	0.799	1	430
OC02	Left	0.073	0.062	1	466
	Right	0.085	0.038	1	463
OC03	Left	0.039	0.185	1	485
	Right	0.053	0.123	1	484
FM01	Left	0.000	0.922	1	480
	Right	0.000	0.871	1	477
FM02	Left	0.000	0.726	1	455
	Right	0.000	0.690	1	457
FM03	Left	0.000	0.403	1	448
	Right	0.020	0.276	1	448
FM04	Left	0.047	0.158	1	456
	Right	0.046	0.162	1	460
FM05	Left	0.000	0.420	1	476
	Right	0.058	0.107	1	473
TB01	Left	0.000	0.451	1	474
	Right	0.000	0.611	1	473
PT01	Left	0.000	0.986	1	442
	Right	0.000	0.353	1	439
PT02	Left	0.127	0.006	1	411
	Right	0.128	0.005	1	413
CLN01	Left	0.044	0.175	1	443
	Right	0.054	0.133	1	436
CLV01	Left	0.000	0.486	2	464
	Right	0.009	0.361	2	468
CLV02	Left	0.000	0.821	1	418
	Right	0.000	0.740	1	435
RB101	Left	0.000	0.551	2	424
	Right	0.000	0.851	2	410
RB102	Left	0.000	0.836	1	408
	Right	0.000	0.660	1	411
PSY01	Left	0.000	0.395	2	400
	Right	0.000	0.740	2	405
PSY02	Left	0.050	0.222	2	402
	Right	0.025	0.323	2	410
PSY03	Left	0.068	0.149	2	394
	Right	0.087	0.082	2	400
IAS01	Left	0.029	0.303	2	459
	Right	0.050	0.205	2	463
IAS02	Left	0.000	0.424	1	469
	Right	0.040	0.186	1	470
SAS01	Left	0.000	0.866	1	430
	Right	0.000	0.672	1	432
SAS02	Left	0.029	0.245	1	424
	Right	0.056	0.124	1	430
ACT01	Left	0.078	0.084	2	483
	Right	0.035	0.275	2	475
ACT02	Left	0.073	0.106	2	466
	Right	0.034	0.283	2	456
ACT03	Left	0.074	0.118	2	417
	Right	0.107	0.034	2	413

Table A.2 Cramér's V for assessment skeletal morphology and sex association on cranial and palatine sutures, and axial traits.

Trait	Statistic	p-value	df	n
CRS01	0.342	0.000	1	457
CRS02L	0.237	0.000	1	454
CRS02R	0.234	0.000	1	453
CRS03	0.065	0.082	1	482
CRS04L	0.054	0.123	1	482
CRS04R	0.070	0.066	1	482
CRS05	0.128	0.003	1	464
CRS06L	0.000	0.862	1	466
CRS06R	0.000	0.718	1	464
C3IS	0.000	0.912	2	446
C4SS	0.000	0.788	2	453
C4IS	0.000	0.427	2	452
C5SS	0.044	0.234	2	459
C5IS	0.000	0.588	2	463
C6SS	0.000	0.573	2	468
C6IS	0.000	0.918	2	466
C7SS	0.041	0.248	2	466
L1IS	0.000	0.631	2	451
L2SS	0.000	0.927	2	454
L2IS	0.000	0.666	2	452
L3SS	0.000	0.848	2	458
L3IS	0.000	0.935	2	457
L4SS	0.000	0.547	2	459
L4IS	0.000	0.696	2	462
L5SS	0.030	0.298	2	457
S1SS	0.043	0.245	2	446
S1S2F	0.102	0.015	1	470

Table A.3 Sex-related differences assessment on conditional age-at-death distribution for bilateral joint and musculoskeletal degenerative traits.

		Left				Right			
Trait	Stage	D	p-value	n(♀)	n(♂)	D	p-value	n(♀)	n(♂)
SC01	0	0.082	0.875	102	108	0.086	0.824	102	110
	1	0.119	0.573	141	132	0.142	0.246	143	133
HM01	0	0.100	0.784	115	127	0.143	0.339	119	121
	1	0.085	0.784	127	110	0.085	0.794	119	115
HM02	0	0.078	0.867	124	113	0.098	0.676	116	103
	1	0.190	0.056	116	122	0.198	0.030	117	133
HM03	0	0.084	0.708	137	145	0.056	0.990	126	126
	1	0.165	0.349	101	81	0.182	0.141	106	96
HM04	0	0.088	0.454	192	188	0.110	0.222	188	173
	1	0.268	0.129	45	51	0.274	0.074	47	61
HM05	0	0.043	0.999	147	152	0.050	0.993	144	147
	1	0.122	0.999	75	64	0.210	0.176	74	68
HM06	0	0.035	1.000	133	141	0.055	0.991	124	131
	1	0.175	0.327	94	72	0.099	0.991	95	81
UL01	0	0.042	0.999	147	166	0.090	0.582	138	163
	1	0.188	0.236	91	71	0.193	0.166	101	74
UL02	0	0.103	0.467	144	130	0.094	0.600	144	124
	1	0.165	0.369	81	94	0.196	0.142	80	95
RD01	0	0.072	0.704	188	194	0.086	0.828	187	193
	1	0.218	0.543	44	40	0.136	0.828	45	40
RD02	0	0.071	0.956	102	105	0.061	0.991	105	99
	1	0.123	0.526	134	135	0.137	0.338	130	134
OC01	0	0.140	0.206	122	110	0.143	0.193	119	109
	1	0.224	0.025	98	103	0.178	0.164	102	100
OC02	0	0.117	0.885	84	109	0.116	0.747	84	110
	1	0.071	0.885	144	129	0.083	0.747	144	125
OC03	0	0.044	1.000	104	120	0.043	1.000	108	125
	1	0.119	0.626	138	123	0.122	0.618	135	116
FM01	0	0.169	0.040	161	163	0.142	0.163	158	160
	1	0.162	0.260	79	77	0.131	0.507	81	78
FM02	0	0.119	0.366	172	167	0.124	0.276	180	167
	1	0.130	0.714	56	60	0.169	0.410	54	56
FM03	0	0.065	0.954	125	128	0.045	0.999	123	132
	1	0.116	0.954	105	90	0.151	0.442	104	89
FM04	0	0.150	0.147	124	110	0.146	0.177	123	107
	1	0.202	0.044	102	120	0.213	0.023	107	123
FM05	0	0.064	0.936	133	145	0.067	0.912	130	150
	1	0.124	0.862	103	95	0.082	0.912	105	88
TB01	0	0.094	0.892	150	155	0.100	0.590	153	156
	1	0.089	0.892	90	79	0.121	0.590	86	78
PT01	0	0.107	0.388	145	141	0.069	0.902	135	141
	1	0.203	0.160	80	76	0.176	0.329	88	75
PT02	0	0.132	0.509	106	134	0.125	0.357	109	136
	1	0.101	0.787	100	71	0.145	0.357	99	69
CLN01	0	0.092	0.734	101	126	0.080	0.869	101	126
	1	0.106	0.734	111	105	0.106	0.869	109	100

Table A.4 Sex-related differences assessment on conditional age-at-death distribution for bilateral *standard* traits.

		Left				Right			
Trait	Stage	D	p-value	n(♀)	n(♂)	D	p-value	n(♀)	n(♂)
CLV01	0	0.231	0.789	25	34	0.228	0.836	25	34
	1	0.073	0.981	82	82	0.071	0.983	84	84
	2	0.105	0.789	123	118	0.102	0.836	127	114
CLV02	0	0.094	0.873	91	97	0.083	0.904	91	95
	1	0.078	0.873	115	115	0.157	0.185	127	122
RB101	0	0.171	0.719	35	31	0.116	0.980	34	32
	1	0.159	0.268	125	121	0.178	0.146	121	115
	2	0.150	0.719	51	61	0.223	0.208	52	56
RB102	0	0.127	0.487	91	82	0.120	0.635	91	81
	1	0.109	0.487	120	115	0.096	0.635	120	119
PSY01	0	0.192	0.591	31	34	0.227	0.902	29	33
	1	0.121	0.568	135	117	0.072	0.902	132	122
	2	0.201	0.568	38	45	0.162	0.902	44	45
PSY02	0	0.249	0.372	40	51	0.224	0.684	38	50
	1	0.077	0.870	132	112	0.055	0.992	131	119
	2	0.217	0.616	32	35	0.152	0.992	35	37
PSY03	0	0.170	0.902	31	30	0.197	0.929	29	30
	1	0.103	0.902	82	96	0.121	0.929	80	99
	2	0.092	0.902	88	67	0.052	1.000	92	70
IAS01	0	0.128	0.742	60	54	0.106	0.914	58	53
	1	0.164	0.551	58	69	0.188	0.364	54	67
	2	0.178	0.192	118	100	0.160	0.316	126	105
IAS02	0	0.059	0.996	94	99	0.039	1.000	90	101
	1	0.161	0.114	146	130	0.170	0.072	150	129
SAS01	0	0.095	0.512	150	146	0.112	0.317	147	148
	1	0.212	0.196	66	68	0.258	0.042	72	65
SAS02	0	0.064	0.984	97	109	0.065	0.980	98	112
	1	0.089	0.984	116	102	0.140	0.471	120	100
ACT01	0	0.123	0.447	107	110	0.129	0.432	106	108
	1	0.273	0.008	98	81	0.180	0.288	102	87
	2	0.188	0.447	35	52	0.208	0.432	31	41
ACT02	0	0.110	0.784	70	72	0.085	0.961	68	73
	1	0.193	0.267	77	56	0.164	0.595	70	52
	2	0.232	0.036	88	103	0.214	0.073	96	97
ACT03	0	0.067	0.998	73	92	0.067	0.993	73	92
	1	0.062	0.998	75	83	0.069	0.993	79	81
	2	0.116	0.998	54	40	0.168	0.993	54	34

Table A.5 Sex-related differences assessment on conditional age-at-death distribution for cranial and palatine suture traits.

Trait	Stage	D	p-value	n(♀)	n(♂)
CRS01	0	0.491	0.000	102	32
	1	0.203	0.004	122	201
CRS02L	0	0.124	0.243	199	154
	1	0.231	0.243	27	74
CRS02R	0	0.129	0.225	199	154
	1	0.201	0.404	27	73
CRS03	0	0.119	0.384	129	106
	1	0.202	0.027	115	132
CRS04L	0	0.219	0.065	96	78
	1	0.115	0.263	146	162
CRS04R	0	0.225	0.057	96	75
	1	0.140	0.098	146	165
CRS05	0	0.169	0.164	106	75
	1	0.154	0.143	125	158
CRS06L	0	0.092	0.633	135	131
	1	0.145	0.490	99	101
CRS06R	0	0.103	0.479	136	130
	1	0.121	0.479	97	101

Table A.6 Sex-related differences assessment on conditional age-at-death distribution for vertebrae traits.

Trait	Stage	D	p-value	n(♀)	n(♂)
C3IS	0	0.094	0.747	103	107
	1	0.140	0.699	74	73
	2	0.288	0.150	46	43
C4SS	0	0.074	0.949	96	104
	1	0.091	0.949	80	79
	2	0.342	0.025	49	45
C4IS	0	0.040	1.000	92	106
	1	0.170	0.346	77	73
	2	0.238	0.320	56	48
C5SS	0	0.060	0.996	85	104
	1	0.127	0.936	77	65
	2	0.196	0.509	65	63
C5IS	0	0.060	0.997	85	97
	1	0.113	0.997	60	56
	2	0.191	0.295	85	80
C6SS	0	0.077	0.947	87	98
	1	0.145	0.909	55	55
	2	0.177	0.397	91	82
C6IS	0	0.129	0.618	92	98
	1	0.136	0.780	47	46
	2	0.201	0.148	92	91
C7SS	0	0.070	0.974	88	100
	1	0.167	0.697	60	46
	2	0.176	0.414	83	89
L1IS	0	0.152	0.188	106	98
	1	0.207	0.188	62	71
	2	0.211	0.188	57	57
L2SS	0	0.131	0.558	97	99
	1	0.078	0.992	60	63
	2	0.199	0.415	69	66
L2IS	0	0.169	0.221	101	95
	1	0.139	0.598	57	66
	2	0.198	0.221	66	67
L3SS	0	0.103	0.744	83	92
	1	0.119	0.744	67	67
	2	0.159	0.744	75	74
L3IS	0	0.117	0.537	94	97
	1	0.183	0.537	64	69
	2	0.140	0.537	67	66
L4SS	0	0.084	0.990	78	91
	1	0.080	0.990	63	58
	2	0.153	0.838	86	83
L4IS	0	0.125	0.675	93	98
	1	0.116	0.846	53	59
	2	0.137	0.675	83	76
L5SS	0	0.065	0.994	80	94
	1	0.134	0.994	51	56
	2	0.115	0.994	95	81
S1SS	0	0.084	0.913	84	95
	1	0.101	0.913	63	71
	2	0.148	0.913	74	59
S1S2F	0	0.224	0.436	26	46
	1	0.087	0.436	209	189

Bilateral asymmetry inferential analysis

Table A.7 Bilateral asymmetry analysis.

Trait	Asymmetry	Left	Right	Statistic	p-value	df	n
SC01	0.034	0.019	0.015	0.250	0.617	1	478
HM01	0.039	0.024	0.015	0.889	0.346	1	465
HM02	0.100	0.028	0.072	8.696	0.003	1	458
HM03	0.133	0.038	0.095	10.593	0.001	1	444
HM04	0.102	0.033	0.070	6.149	0.013	1	460
HM05	0.098	0.042	0.056	0.900	0.343	1	410
HM06	0.078	0.020	0.059	8.000	0.005	1	410
UL01	0.086	0.030	0.056	3.600	0.058	1	464
UL02	0.065	0.023	0.042	2.286	0.131	1	428
RD01	0.096	0.045	0.051	0.209	0.647	1	448
RD02	0.046	0.020	0.026	0.429	0.513	1	455
OC01	0.024	0.007	0.017	1.600	0.206	1	410
OC02	0.025	0.013	0.011	0.091	0.763	1	449
OC03	0.036	0.027	0.008	4.765	0.029	1	475
FM01	0.075	0.032	0.043	0.714	0.398	1	465
FM02	0.056	0.033	0.023	0.667	0.414	1	431
FM03	0.040	0.021	0.019	0.059	0.808	1	429
FM04	0.044	0.018	0.025	0.474	0.491	1	437
FM05	0.075	0.041	0.035	0.257	0.612	1	464
TB01	0.069	0.039	0.030	0.500	0.480	1	462
PT01	0.059	0.019	0.040	3.240	0.072	1	422
PT02	0.028	0.018	0.010	0.818	0.366	1	395
CLN01	0.019	0.012	0.007	0.500	0.480	1	428
CLV01	0.002	0.002	0.000	1.000	0.317	1	448
CLV02	0.068	0.030	0.038	0.333	0.564	1	400
RB101	0.019	0.008	0.011	0.143	0.706	1	373
RB102	0.008	0.003	0.005	0.333	0.564	1	367
PSY01	0.011	0.008	0.003	1.000	0.317	1	368
PSY02	0.016	0.008	0.008	0.000	1.000	1	367
PSY03	0.011	0.003	0.008	1.000	0.317	1	361
IAS01	0.034	0.007	0.027	5.400	0.020	1	441
IAS02	0.018	0.009	0.009	0.000	1.000	1	454
SAS01	0.010	0.002	0.007	1.000	0.317	1	411
SAS02	0.005	0.002	0.002	0.000	1.000	1	410
ACT01	0.045	0.026	0.019	0.429	0.513	1	468
ACT02	0.061	0.014	0.047	8.333	0.004	1	445
ACT03	0.010	0.008	0.003	1.000	0.317	1	397

Table A.8 Sex-specific asymmetry analysis for joint and musculoskeletal degenerative traits.

Trait	Sex	Asymmetry	Left	Right	Statistic	p-value	df	n
SC01	Female	0.025	0.012	0.012	0.000	1.000	1	242
	Male	0.042	0.025	0.017	0.400	0.527	1	236
HM01	Female	0.038	0.030	0.009	2.778	0.096	1	235
	Male	0.039	0.017	0.022	0.111	0.739	1	230
HM02	Female	0.088	0.026	0.061	3.200	0.074	1	228
	Male	0.113	0.030	0.083	5.539	0.019	1	230
HM03	Female	0.106	0.031	0.075	4.167	0.041	1	227
	Male	0.161	0.046	0.115	6.429	0.011	1	217
HM04	Female	0.096	0.035	0.061	1.636	0.201	1	230
	Male	0.109	0.030	0.078	4.840	0.028	1	230
HM05	Female	0.073	0.029	0.044	0.600	0.439	1	207
	Male	0.123	0.054	0.069	0.360	0.549	1	203
HM06	Female	0.081	0.019	0.062	4.765	0.029	1	211
	Male	0.075	0.020	0.055	3.267	0.071	1	199
UL01	Female	0.086	0.022	0.065	5.000	0.025	1	232
	Male	0.086	0.039	0.047	0.200	0.655	1	232
UL02	Female	0.028	0.014	0.014	0.000	1.000	1	216
	Male	0.104	0.033	0.071	2.909	0.088	1	212
RD01	Female	0.108	0.049	0.058	0.167	0.683	1	223
	Male	0.084	0.040	0.044	0.053	0.819	1	225
RD02	Female	0.044	0.018	0.027	0.400	0.527	1	226
	Male	0.048	0.022	0.026	0.091	0.763	1	229
OC01	Female	0.024	0.005	0.019	1.800	0.180	1	212
	Male	0.025	0.010	0.015	0.200	0.655	1	198
OC02	Female	0.018	0.009	0.009	0.000	1.000	1	220
	Male	0.031	0.018	0.013	0.143	0.706	1	229
OC03	Female	0.017	0.017	0.000	4.000	0.046	1	239
	Male	0.055	0.038	0.017	1.923	0.166	1	236
FM01	Female	0.090	0.039	0.051	0.429	0.513	1	234
	Male	0.061	0.026	0.035	0.286	0.593	1	231
FM02	Female	0.055	0.037	0.018	1.333	0.248	1	219
	Male	0.057	0.028	0.028	0.000	1.000	1	212
FM03	Female	0.032	0.014	0.018	0.143	0.706	1	221
	Male	0.048	0.029	0.019	0.400	0.527	1	208
FM04	Female	0.037	0.014	0.023	0.500	0.480	1	218
	Male	0.050	0.023	0.027	0.091	0.763	1	219
FM05	Female	0.078	0.030	0.048	0.889	0.346	1	230
	Male	0.073	0.051	0.021	2.882	0.090	1	234
TB01	Female	0.089	0.047	0.042	0.048	0.827	1	236
	Male	0.049	0.031	0.018	0.818	0.366	1	226
PT01	Female	0.083	0.023	0.060	3.556	0.059	1	217
	Male	0.034	0.015	0.020	0.143	0.706	1	205
PT02	Female	0.010	0.010	0.000	2.000	0.157	1	199
	Male	0.046	0.026	0.020	0.111	0.739	1	196
CLN01	Female	0.010	0.005	0.005	0.000	1.000	1	206
	Male	0.027	0.018	0.009	0.667	0.414	1	222

Table A.9 Sex-specific asymmetry analysis for *standard* skeletal traits.

Trait	Sex	Asymmetry	Left	Right	Statistic	p-value	df	n
CLV01	Female	0.000	0.000	0.000	0.000	1.000	1	225
	Male	0.005	0.005	0.000	1.000	0.317	1	223
CLV02	Female	0.060	0.035	0.025	0.333	0.564	1	200
	Male	0.075	0.025	0.050	1.667	0.197	1	200
RB101	Female	0.021	0.005	0.016	1.000	0.317	1	188
	Male	0.016	0.011	0.005	0.333	0.564	1	185
RB102	Female	0.005	0.000	0.005	1.000	0.317	1	192
	Male	0.011	0.006	0.006	0.000	1.000	1	175
PSY01	Female	0.005	0.000	0.005	1.000	0.317	1	189
	Male	0.017	0.017	0.000	3.000	0.083	1	179
PSY02	Female	0.011	0.005	0.005	0.000	1.000	1	187
	Male	0.022	0.011	0.011	0.000	1.000	1	180
PSY03	Female	0.011	0.000	0.011	2.000	0.157	1	185
	Male	0.011	0.006	0.006	0.000	1.000	1	176
IAS01	Female	0.035	0.009	0.026	2.000	0.157	1	229
	Male	0.033	0.005	0.028	3.571	0.059	1	212
IAS02	Female	0.004	0.000	0.004	1.000	0.317	1	233
	Male	0.032	0.018	0.014	0.143	0.706	1	221
SAS01	Female	0.005	0.000	0.005	1.000	0.317	1	207
	Male	0.015	0.005	0.010	0.333	0.564	1	204
SAS02	Female	0.000	0.000	0.000	0.000	1.000	1	207
	Male	0.010	0.005	0.005	0.000	1.000	1	203
ACT01	Female	0.043	0.026	0.017	0.400	0.527	1	234
	Male	0.047	0.026	0.021	0.091	0.763	1	234
ACT02	Female	0.101	0.022	0.079	7.348	0.007	1	228
	Male	0.018	0.005	0.014	1.000	0.317	1	217
ACT03	Female	0.015	0.010	0.005	0.333	0.564	1	196
	Male	0.005	0.005	0.000	1.000	0.317	1	201

Table A.10 Age-related asymmetry analysis for joint and musculoskeletal degenerative traits.

		Pooled			Female			Male		
Trait	Stage	D	p-	n	D	p-	n	D	p-	n
SC01	0	0.019	1.000	209	0.029	1.000	102	0.018	1.000	107
	1	0.010	1.000	269	0.021	1.000	140	0.023	1.000	129
HM01	0	0.033	1.000	235	0.047	0.999	113	0.017	1.000	122
	1	0.027	1.000	230	0.034	1.000	122	0.023	1.000	108
HM02	0	0.052	0.927	234	0.040	1.000	122	0.074	0.937	112
	1	0.040	0.993	224	0.042	1.000	106	0.037	1.000	118
HM03	0	0.069	0.553	275	0.065	0.945	134	0.080	0.790	141
	1	0.037	1.000	169	0.025	1.000	93	0.065	0.995	76
HM04	0	0.025	1.000	370	0.021	1.000	190	0.037	1.000	180
	1	0.049	1.000	90	0.055	1.000	40	0.077	0.997	50
HM05	0	0.012	1.000	287	0.023	1.000	143	0.016	1.000	144
	1	0.043	1.000	123	0.050	1.000	64	0.058	1.000	59
HM06	0	0.046	0.945	265	0.057	0.985	132	0.036	1.000	133
	1	0.073	0.805	145	0.098	0.821	79	0.049	1.000	66
UL01	0	0.029	0.999	308	0.057	0.977	145	0.020	1.000	163
	1	0.028	1.000	156	0.036	1.000	87	0.052	1.000	69
UL02	0	0.019	1.000	266	0.014	1.000	140	0.042	1.000	126
	1	0.023	1.000	162	0.026	1.000	76	0.024	1.000	86
RD01	0	0.016	1.000	368	0.014	1.000	181	0.021	1.000	187
	1	0.059	0.999	80	0.044	1.000	42	0.095	0.995	38
RD02	0	0.010	1.000	201	0.038	1.000	102	0.033	1.000	99
	1	0.011	1.000	254	0.021	1.000	124	0.030	1.000	130
OC01	0	0.020	1.000	228	0.019	1.000	121	0.028	1.000	107
	1	0.015	1.000	182	0.027	1.000	91	0.026	1.000	91
OC02	0	0.017	1.000	192	0.012	1.000	84	0.021	1.000	108
	1	0.015	1.000	257	0.007	1.000	136	0.024	1.000	121
OC03	0	0.030	1.000	224	0.033	1.000	104	0.031	1.000	120
	1	0.016	1.000	251	0.011	1.000	135	0.031	1.000	116
FM01	0	0.011	1.000	319	0.028	1.000	159	0.020	1.000	160
	1	0.023	1.000	146	0.045	1.000	75	0.060	0.999	71
FM02	0	0.013	1.000	332	0.015	1.000	169	0.012	1.000	163
	1	0.029	1.000	99	0.058	1.000	50	0.041	1.000	49
FM03	0	0.015	1.000	252	0.018	1.000	124	0.015	1.000	128
	1	0.018	1.000	177	0.030	1.000	97	0.014	1.000	80
FM04	0	0.014	1.000	232	0.017	1.000	124	0.019	1.000	108
	1	0.025	1.000	205	0.033	1.000	94	0.024	1.000	111
FM05	0	0.015	1.000	274	0.035	1.000	132	0.030	1.000	142
	1	0.025	1.000	190	0.037	1.000	98	0.077	0.955	92
TB01	0	0.012	1.000	300	0.009	1.000	149	0.020	1.000	151
	1	0.031	1.000	162	0.022	1.000	87	0.044	1.000	75
PT01	0	0.022	1.000	278	0.037	1.000	141	0.007	1.000	137
	1	0.021	1.000	144	0.047	1.000	76	0.013	1.000	68
PT02	0	0.024	1.000	238	0.018	1.000	106	0.030	1.000	132
	1	0.027	1.000	157	0.011	1.000	93	0.059	1.000	64
CLN01	0	0.009	1.000	224	0.010	1.000	100	0.010	1.000	124
	1	0.007	1.000	204	0.009	1.000	106	0.025	1.000	98

Table A.11 Age-related asymmetry analysis for *standard* skeletal traits.

		Pooled			Female			Male		
Trait	Stage	D	p-value	n	D	p-value	n	D	p-value	n
CLV01	0	0.000	1.000	58	0.000	1.000	25	0.000	1.000	33
	1	0.006	1.000	163	0.000	1.000	82	0.012	1.000	81
	2	0.003	1.000	227	0.000	1.000	118	0.006	1.000	109
CLV02	0	0.017	1.000	179	0.032	1.000	88	0.035	1.000	91
	1	0.020	1.000	221	0.028	1.000	112	0.033	1.000	109
RB101	0	0.016	1.000	63	0.000	1.000	34	0.033	1.000	29
	1	0.009	1.000	218	0.017	1.000	111	0.009	1.000	107
	2	0.016	1.000	92	0.031	1.000	43	0.020	1.000	49
RB102	0	0.011	1.000	166	0.009	1.000	90	0.013	1.000	76
	1	0.006	1.000	201	0.009	1.000	102	0.010	1.000	99
PSY01	0	0.000	1.000	61	0.000	1.000	29	0.000	1.000	32
	1	0.005	1.000	231	0.007	1.000	125	0.018	1.000	106
	2	0.025	1.000	76	0.014	1.000	35	0.027	1.000	41
PSY02	0	0.024	1.000	84	0.027	1.000	37	0.021	1.000	47
	1	0.009	1.000	224	0.008	1.000	123	0.010	1.000	101
	2	0.017	1.000	59	0.000	1.000	27	0.031	1.000	32
PSY03	0	0.018	1.000	56	0.000	1.000	28	0.036	1.000	28
	1	0.011	1.000	166	0.020	1.000	79	0.011	1.000	87
	2	0.004	1.000	139	0.007	1.000	78	0.000	1.000	61
IAS01	0	0.018	1.000	112	0.017	1.000	59	0.027	1.000	53
	1	0.047	1.000	120	0.035	1.000	55	0.058	1.000	65
	2	0.012	1.000	209	0.021	1.000	115	0.017	1.000	94
IAS02	0	0.011	1.000	189	0.010	1.000	91	0.015	1.000	98
	1	0.008	1.000	265	0.006	1.000	142	0.016	1.000	123
SAS01	0	0.010	1.000	289	0.007	1.000	145	0.014	1.000	144
	1	0.018	1.000	122	0.015	1.000	62	0.028	1.000	60
SAS02	0	0.005	1.000	205	0.000	1.000	96	0.009	1.000	109
	1	0.005	1.000	205	0.000	1.000	111	0.011	1.000	94
ACT01	0	0.010	1.000	217	0.011	1.000	107	0.017	1.000	110
	1	0.011	1.000	175	0.031	1.000	95	0.051	1.000	80
	2	0.029	1.000	76	0.078	1.000	32	0.087	0.997	44
ACT02	0	0.013	1.000	142	0.024	1.000	70	0.009	1.000	72
	1	0.038	1.000	131	0.060	0.999	77	0.021	1.000	54
	2	0.044	0.995	172	0.084	0.921	81	0.019	1.000	91
ACT03	0	0.006	1.000	165	0.014	1.000	73	0.000	1.000	92
	1	0.011	1.000	152	0.015	1.000	74	0.011	1.000	78
	2	0.011	1.000	80	0.011	1.000	49	0.019	1.000	31

Missing data analysis

Table A.12 Percentage of missing values by skeletal trait (n=500)

Trait	Missing	Trait	Left	Right	Merged	Trait	Left	Right	Merged
CRS01	8.60%	CLV01	7.20%	6.40%	3.20%	SC01	3.40%	2.40%	1.40%
CRS02L	9.20%	CLV02	16.40%	13.00%	9.40%	HM01	4.20%	5.20%	2.40%
CRS02R	9.40%					HM02	5.00%	6.20%	2.80%
CRS03	3.60%	RB101	15.20%	18.00%	7.80%	HM03	7.20%	9.20%	5.20%
CRS04L	3.60%	RB102	18.40%	17.80%	9.60%	HM04	4.80%	6.20%	3.00%
CRS04R	3.60%					HM05	12.40%	13.40%	7.80%
CRS05	7.20%	PSY01	20.00%	19.00%	12.60%	HM06	12.00%	13.80%	7.80%
CRS06L	6.80%	PSY02	19.60%	18.00%	11.00%	UL01	5.00%	4.80%	2.60%
CRS06R	7.20%	PSY03	21.20%	20.00%	13.40%	UL02	10.20%	11.40%	7.20%
						RD01	6.80%	7.00%	3.40%
C3IS	10.80%	IAS01	8.20%	7.40%	3.80%	RD02	4.80%	6.40%	2.20%
C4SS	9.40%	IAS02	6.20%	6.00%	3.00%				
C4IS	9.60%	SAS01	14.00%	13.60%	9.80%	OC01	13.40%	14.00%	9.40%
C5SS	8.20%	SAS02	15.20%	14.00%	11.20%	OC02	6.80%	7.40%	4.00%
C5IS	7.40%					OC03	3.00%	3.20%	1.20%
C6SS	6.40%	ACT01	3.40%	5.00%	2.00%	FM01	4.00%	4.60%	1.60%
C6IS	6.80%	ACT02	6.80%	8.80%	4.60%	FM02	9.00%	8.60%	3.80%
C7SS	6.80%	ACT03	16.60%	17.40%	13.40%	FM03	10.40%	10.40%	6.60%
L1IS	9.80%					FM04	8.80%	8.00%	4.20%
L2SS	9.20%					FM05	4.80%	5.40%	3.00%
L2IS	9.60%					TB01	5.20%	5.40%	3.00%
L3SS	8.40%					PT01	11.60%	12.20%	8.20%
L3IS	8.60%					PT02	17.80%	17.40%	14.20%
L4SS	8.20%					CLN01	11.40%	12.80%	9.80%
L4IS	7.60%								
L5SS	8.60%								
S1SS	10.80%								
S1S2F	6.00%								

Appendix B

Age-related descriptive statistics

Table B.1 Descriptive age-related statistics for joint and musculoskeletal degenerative traits. Left side.

		Female					Male				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
SC01	0	36.608	12.893	19.000	66.900	102	35.926	13.322	19.000	67.950	108
	1	74.745	14.844	47.000	97.500	141	70.826	14.958	41.275	92.725	132
HM01	0	42.287	19.883	19.000	88.000	115	38.843	14.843	19.150	73.550	127
	1	73.386	15.211	46.000	97.000	127	72.691	14.057	42.725	93.275	110
HM02	0	41.395	16.259	19.000	76.925	124	40.469	16.755	19.800	77.400	113
	1	77.164	14.198	49.125	98.125	116	70.139	16.324	32.075	92.000	122
HM03	0	45.161	19.122	19.000	85.000	137	47.724	20.370	20.000	88.200	145
	1	76.386	15.673	42.500	98.500	101	71.284	16.122	42.000	92.000	81
HM04	0	53.224	22.567	19.000	95.000	192	49.447	20.990	20.000	89.325	188
	1	79.822	12.132	50.800	97.000	45	74.216	14.084	49.250	95.000	51
HM05	0	45.238	18.312	19.000	84.400	147	45.138	18.599	19.775	84.225	152
	1	78.893	13.070	51.700	99.150	75	76.313	12.164	54.000	92.425	64
HM06	0	42.301	16.368	19.000	75.700	133	43.440	17.601	19.500	88.000	141
	1	78.830	12.841	52.325	98.675	94	75.528	11.410	54.775	94.450	72
UL01	0	47.204	20.940	19.000	92.700	147	47.000	20.514	20.000	89.000	166
	1	76.923	13.712	50.500	98.750	91	72.915	14.303	46.000	93.750	71
UL02	0	47.694	21.997	19.000	92.425	144	43.854	19.783	20.000	85.000	130
	1	74.395	13.822	50.000	97.000	81	68.915	16.426	34.600	92.000	94
RD01	0	53.255	22.842	19.000	94.325	188	50.716	21.507	20.000	90.350	194
	1	77.205	14.692	52.000	97.000	44	73.950	14.459	48.475	93.075	40
RD02	0	38.235	15.938	19.000	77.950	102	36.600	13.993	19.000	69.800	105
	1	73.575	15.383	45.325	97.675	134	69.030	16.597	32.350	92.650	135
OC01	0	41.107	16.762	19.000	79.975	122	37.618	15.292	19.000	71.825	110
	1	75.133	13.983	50.850	97.575	98	66.612	16.324	32.550	93.450	103
OC02	0	34.845	14.271	19.000	65.925	84	37.734	15.090	19.000	76.900	109
	1	69.847	16.523	40.725	97.000	144	68.333	16.325	32.400	92.800	129
OC03	0	36.865	13.377	19.000	67.700	104	37.483	14.007	19.000	71.075	120
	1	74.877	14.232	49.425	97.575	138	71.228	14.214	44.100	92.950	123
FM01	0	50.578	22.847	19.000	92.000	161	43.939	17.644	20.000	81.950	163
	1	73.937	14.803	51.800	99.050	79	76.195	11.764	54.800	94.200	77
FM02	0	50.948	22.296	19.000	94.725	172	46.503	19.988	20.000	89.850	167
	1	75.786	13.145	53.125	96.875	56	73.917	12.920	45.850	93.050	60
FM03	0	41.072	16.120	19.000	73.900	125	39.227	14.337	19.175	70.125	128
	1	76.019	13.367	52.600	98.400	105	72.667	13.990	44.350	93.775	90
FM04	0	40.935	15.837	19.000	74.000	124	36.155	12.740	19.000	64.550	110
	1	76.167	13.732	52.000	97.475	102	69.450	15.379	33.975	93.025	120
FM05	0	44.752	19.876	19.000	88.700	133	42.448	18.011	19.600	85.400	145
	1	75.369	14.572	45.550	98.450	103	72.663	13.784	46.400	93.650	95
TB01	0	48.433	21.910	19.000	92.825	150	44.387	19.048	19.850	87.000	155
	1	74.967	14.731	46.675	97.775	90	72.987	14.271	47.400	94.100	79
PT01	0	45.862	19.570	19.000	87.400	145	42.574	18.326	19.500	85.500	141
	1	75.900	14.225	48.925	97.025	80	70.408	13.384	47.625	93.125	76
PT02	0	37.453	13.797	19.000	68.375	106	41.888	17.643	19.325	81.350	134
	1	73.010	14.701	46.950	98.000	100	70.915	14.829	42.000	92.500	71
CLN01	0	37.446	15.464	19.000	84.000	101	39.127	15.683	19.125	75.000	126
	1	73.063	14.407	48.250	97.250	111	71.876	14.401	41.600	92.000	105

Table B.2 Descriptive age-related statistics for joint and musculoskeletal degenerative traits. Right side.

		Female					Male				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
SC01	0	37.373	13.729	19.000	69.475	102	36.418	13.882	19.000	74.550	110
	1	74.252	15.594	43.000	97.450	143	70.647	14.351	43.000	92.700	133
HM01	0	44.000	21.562	19.000	92.200	119	38.198	14.340	19.000	71.000	121
	1	71.992	14.778	46.000	97.000	119	72.443	13.663	42.850	93.150	115
HM02	0	39.681	14.962	19.000	69.625	116	38.485	15.602	19.550	74.250	103
	1	75.393	14.710	43.000	97.100	117	68.504	16.579	32.600	92.700	133
HM03	0	42.627	17.613	19.000	81.375	126	44.429	19.292	20.000	86.750	126
	1	75.208	15.177	46.875	97.375	106	70.156	15.528	42.375	92.000	96
HM04	0	52.590	22.342	19.000	94.000	188	48.694	21.042	20.000	89.000	173
	1	79.298	12.729	48.650	97.850	47	72.656	14.299	49.500	94.000	61
HM05	0	44.618	18.025	19.000	82.700	144	45.639	19.392	19.650	85.700	147
	1	79.770	13.138	51.650	97.175	74	74.456	12.742	50.675	92.325	68
HM06	0	40.919	16.244	19.000	78.850	124	41.580	16.560	19.250	78.500	131
	1	76.074	13.431	52.350	97.650	95	74.506	11.751	54.000	93.000	81
UL01	0	44.862	19.341	19.000	87.000	138	46.712	20.630	20.000	88.950	163
	1	77.168	14.279	48.000	97.500	101	72.649	13.940	48.825	93.525	74
UL02	0	47.139	21.387	19.000	90.275	144	43.347	18.730	20.000	80.925	124
	1	74.350	13.854	50.000	97.075	80	67.463	17.269	29.350	92.000	95
RD01	0	53.299	22.984	19.000	94.350	187	50.244	20.991	20.000	89.000	193
	1	76.889	15.280	49.300	98.800	45	76.775	13.383	52.900	96.000	40
RD02	0	39.038	17.102	19.000	87.000	105	37.899	15.072	19.450	70.100	99
	1	73.869	15.088	46.225	97.775	130	68.537	16.883	32.325	92.675	134
OC01	0	40.756	16.883	19.000	82.050	119	36.266	13.362	19.000	66.900	109
	1	74.510	14.105	50.000	97.000	102	67.760	15.683	34.850	92.525	100
OC02	0	34.548	13.924	19.000	64.775	84	36.809	14.250	19.000	73.375	110
	1	69.951	16.174	42.575	96.425	144	68.592	15.608	34.400	92.900	125
OC03	0	38.130	14.961	19.000	71.300	108	38.376	15.094	19.100	75.800	125
	1	75.007	14.087	49.350	97.000	135	72.078	13.766	43.875	93.125	116
FM01	0	49.886	22.443	19.000	94.150	158	43.881	18.031	19.975	82.075	160
	1	74.037	14.735	48.000	96.000	81	75.705	12.085	55.000	94.150	78
FM02	0	51.706	22.166	19.000	93.050	180	46.449	20.205	20.000	91.700	167
	1	77.204	13.102	52.975	98.375	54	72.518	14.054	42.375	91.250	56
FM03	0	40.724	15.623	19.000	73.000	123	39.985	15.124	19.275	74.175	132
	1	75.510	13.818	51.150	97.850	104	72.191	14.530	43.200	93.800	89
FM04	0	41.398	16.938	19.000	76.950	123	36.299	13.342	19.000	72.400	107
	1	76.224	13.844	52.000	97.000	107	69.252	15.345	40.050	92.950	123
FM05	0	44.208	19.827	19.000	88.550	130	42.453	17.192	19.725	80.550	150
	1	74.943	14.312	48.600	97.000	105	74.364	13.255	49.000	93.825	88
TB01	0	48.948	22.293	19.000	95.200	153	44.333	18.748	19.875	87.000	156
	1	75.279	14.507	49.125	97.000	86	73.359	14.048	45.250	94.150	78
PT01	0	44.585	19.535	19.000	88.650	135	42.986	18.674	19.500	86.000	141
	1	74.477	14.025	49.175	97.000	88	70.133	13.325	47.550	92.300	75
PT02	0	38.633	15.383	19.000	76.200	109	42.816	18.835	19.375	86.250	136
	1	73.404	14.380	47.350	97.550	99	68.609	13.991	41.800	92.000	69
CLN01	0	37.416	15.308	19.000	81.000	101	38.738	15.194	19.125	74.625	126
	1	72.358	14.674	46.000	97.300	109	71.760	14.042	41.475	92.000	100

Table B.3 Descriptive age-related statistics for joint and musculoskeletal degenerative traits. Pooled sexes.

		Left					Right				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
SC01	0	36.257	13.088	19.000	68.325	210	36.877	13.784	19.000	73.450	212
	1	72.850	15.000	43.000	96.200	273	72.514	15.089	43.000	96.125	276
HM01	0	40.479	17.469	19.000	86.925	242	41.075	18.473	19.000	88.000	240
	1	73.063	14.660	43.000	96.100	237	72.214	14.212	45.475	96.000	234
HM02	0	40.954	16.469	19.000	77.000	237	39.119	15.243	19.000	73.100	219
	1	73.563	15.692	37.925	97.000	238	71.728	16.075	37.225	96.000	250
HM03	0	46.479	19.780	19.025	87.000	282	43.528	18.457	19.000	85.000	252
	1	74.115	16.033	42.000	97.000	182	72.807	15.514	43.000	96.975	202
HM04	0	51.355	21.854	19.475	92.525	380	50.723	21.786	19.000	92.000	361
	1	76.844	13.435	49.375	96.625	96	75.546	13.974	48.350	97.000	108
HM05	0	45.187	18.427	19.000	84.550	299	45.134	18.703	19.000	85.000	291
	1	77.705	12.681	52.450	97.550	139	77.225	13.177	50.525	97.000	142
HM06	0	42.887	16.992	19.000	80.400	274	41.259	16.378	19.000	79.000	255
	1	77.398	12.315	54.000	97.875	166	75.352	12.674	53.000	97.000	176
UL01	0	47.096	20.682	19.000	90.400	313	45.864	20.037	19.000	88.500	301
	1	75.167	14.072	48.025	96.975	162	75.257	14.273	48.000	97.000	175
UL02	0	45.872	21.025	19.000	89.525	274	45.384	20.253	19.000	87.000	268
	1	71.451	15.478	43.000	95.000	175	70.611	16.127	35.850	95.650	175
RD01	0	51.966	22.181	19.525	92.000	382	51.747	22.018	19.475	92.000	380
	1	75.655	14.585	49.075	96.925	84	76.835	14.333	49.300	97.000	85
RD02	0	37.406	14.969	19.000	75.700	207	38.485	16.119	19.000	84.325	204
	1	71.294	16.135	37.700	96.300	269	71.163	16.215	37.575	96.425	264
OC01	0	39.453	16.142	19.000	79.225	232	38.610	15.433	19.000	79.325	228
	1	70.766	15.778	40.000	96.000	201	71.168	15.250	42.025	96.000	202
OC02	0	36.477	14.771	19.000	74.400	193	35.830	14.118	19.000	72.350	194
	1	69.132	16.417	37.800	96.000	273	69.320	15.898	38.700	95.300	269
OC03	0	37.196	13.691	19.000	70.425	224	38.262	15.001	19.000	74.400	233
	1	73.157	14.313	48.000	96.500	261	73.653	13.989	48.000	96.000	251
FM01	0	47.238	20.634	19.000	89.000	324	46.865	20.532	19.000	89.225	318
	1	75.051	13.395	52.000	96.125	156	74.855	13.484	51.800	96.000	159
FM02	0	48.758	21.276	19.000	92.000	339	49.176	21.377	19.000	92.000	347
	1	74.819	13.006	49.000	95.125	116	74.818	13.736	47.350	95.000	110
FM03	0	40.138	15.242	19.000	72.700	253	40.341	15.341	19.000	73.650	255
	1	74.472	13.725	50.000	97.000	195	73.979	14.211	48.800	96.200	193
FM04	0	38.688	14.630	19.000	74.000	234	39.026	15.548	19.000	76.000	230
	1	72.536	14.993	41.525	96.000	222	72.496	15.043	42.725	96.000	230
FM05	0	43.550	18.927	19.000	88.000	278	43.268	18.449	19.000	85.050	280
	1	74.071	14.228	45.000	97.000	198	74.679	13.807	48.800	96.000	193
TB01	0	46.377	20.572	19.000	89.000	305	46.618	20.676	19.000	89.600	309
	1	74.041	14.508	46.400	96.800	169	74.366	14.279	46.150	96.000	164
PT01	0	44.241	19.005	19.000	87.000	286	43.768	19.082	19.000	87.125	276
	1	73.224	14.049	47.750	96.125	156	72.479	13.836	48.050	96.000	163
PT02	0	39.929	16.178	19.000	79.025	240	40.955	17.475	19.000	84.000	245
	1	72.140	14.747	45.000	96.750	171	71.435	14.375	45.000	95.825	168
CLN01	0	38.379	15.574	19.000	76.350	227	38.150	15.225	19.000	77.050	227
	1	72.486	14.382	45.000	96.000	216	72.072	14.344	45.000	96.000	209

Table B.4 Descriptive age-related statistics for standard skeletal markers. Left side.

		Female					Male				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
CLV01	0	21.880	2.438	19.000	26.400	25	23.324	3.245	19.000	29.175	34
	1	43.110	12.275	26.000	74.000	82	41.976	10.272	27.025	64.000	82
	2	75.293	13.772	50.100	97.000	123	72.636	12.660	50.000	93.075	118
CLV02	0	36.220	14.627	19.000	73.000	91	37.536	15.612	19.000	73.000	97
	1	70.991	16.270	41.550	97.150	115	68.565	16.848	32.000	93.150	115
RB101	0	25.057	6.131	19.000	40.750	35	23.258	3.473	19.000	30.000	31
	1	56.352	18.062	28.100	91.600	125	49.777	16.303	26.000	83.000	121
	2	80.588	11.810	57.250	97.750	51	77.246	11.679	55.000	95.000	61
RB102	0	36.077	13.373	19.000	71.500	91	32.878	10.697	19.000	54.975	82
	1	71.500	15.472	42.900	96.025	120	67.713	15.697	38.850	93.150	115
PSY01	0	24.258	5.698	19.000	38.000	31	25.029	4.783	19.000	34.525	34
	1	57.674	19.212	27.350	94.600	135	54.333	16.131	29.000	87.000	117
	2	78.605	13.347	54.625	97.075	38	74.556	13.093	52.300	92.900	45
PSY02	0	27.175	9.337	19.000	47.375	40	29.353	8.946	19.000	52.000	51
	1	58.871	18.693	29.000	94.900	132	57.830	16.236	31.550	88.225	112
	2	80.063	12.536	53.875	95.675	32	75.914	11.184	56.700	93.450	35
PSY03	0	24.516	8.213	19.000	44.000	31	25.333	6.697	19.000	40.300	30
	1	49.732	17.152	27.000	86.925	82	51.521	16.647	27.375	85.500	96
	2	71.955	15.784	39.700	97.650	88	70.776	14.153	43.000	93.350	67
IAS01	0	30.417	9.977	19.000	50.575	60	28.759	8.452	19.000	48.675	54
	1	54.017	16.008	30.850	81.000	58	48.754	15.604	26.000	82.400	69
	2	75.195	15.073	51.850	98.000	118	70.770	14.321	41.950	92.525	100
IAS02	0	36.915	15.195	19.000	80.025	94	35.697	13.807	19.000	69.000	99
	1	72.534	15.979	41.250	97.375	146	67.169	16.222	34.450	92.000	130
SAS01	0	46.647	19.177	19.000	85.550	150	43.925	18.517	19.625	84.250	146
	1	77.136	13.530	51.625	96.375	66	71.162	13.585	47.050	92.325	68
SAS02	0	39.278	17.673	19.000	84.200	97	37.982	15.241	19.000	74.900	109
	1	69.103	16.140	39.000	96.000	116	68.098	14.352	39.050	92.000	102
ACT01	0	38.822	15.385	19.000	75.750	107	35.827	13.155	19.000	66.100	110
	1	71.173	15.411	40.700	96.150	98	63.988	13.268	41.000	89.000	81
	2	81.171	12.213	57.250	97.150	35	79.962	10.006	60.825	95.450	52
ACT02	0	32.371	11.263	19.000	60.375	70	30.764	9.226	19.000	49.000	72
	1	56.260	14.684	28.800	84.300	77	50.375	14.597	26.000	78.000	56
	2	79.386	12.808	53.175	97.000	88	72.806	13.516	43.000	92.450	103
ACT03	0	32.123	10.406	19.000	51.800	73	33.011	11.086	19.000	60.000	92
	1	60.547	15.834	32.400	92.000	75	60.651	14.630	36.050	87.000	83
	2	80.074	12.631	53.650	96.675	54	79.150	11.437	52.950	96.000	40

Table B.5 Descriptive age-related statistics for standard skeletal markers. Right side.

		Female					Male				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
CLV01	0	21.880	2.438	19.000	26.400	25	23.235	3.210	19.000	29.175	34
	1	43.655	12.929	26.000	76.775	84	42.440	10.627	27.075	64.925	84
	2	75.606	14.069	50.300	97.850	127	73.149	12.949	50.000	93.175	114
CLV02	0	36.319	13.787	19.000	66.750	91	38.011	16.032	19.000	78.650	95
	1	72.079	16.495	39.450	97.000	127	67.270	16.869	32.150	92.000	122
RB101	0	24.618	5.635	19.000	36.750	34	23.969	5.550	19.000	34.050	32
	1	55.215	17.257	27.000	88.000	121	49.383	15.907	26.000	82.900	115
	2	80.519	12.315	57.275	98.725	52	77.089	10.933	56.125	93.625	56
RB102	0	36.011	13.338	19.000	71.500	91	32.728	10.557	19.000	54.000	81
	1	70.750	15.896	38.975	96.025	120	68.370	14.856	42.000	92.050	119
PSY01	0	23.724	5.257	19.000	38.000	29	25.152	4.777	19.000	34.600	33
	1	55.909	18.380	27.275	89.725	132	54.844	16.460	28.025	87.000	122
	2	78.432	13.932	50.375	96.850	44	76.356	13.023	52.300	92.900	45
PSY02	0	26.447	7.515	19.000	42.375	38	28.720	7.941	19.225	47.875	50
	1	57.878	17.905	29.000	90.750	131	57.731	16.551	28.950	89.000	119
	2	80.114	12.792	54.250	95.450	35	78.108	10.582	59.700	93.300	37
PSY03	0	24.379	8.248	19.000	43.100	29	26.033	7.554	19.000	46.100	30
	1	48.825	16.949	26.975	88.075	80	51.253	17.502	26.450	85.200	99
	2	71.489	15.703	39.000	95.000	92	71.414	14.498	43.000	93.275	70
IAS01	0	29.759	9.336	19.000	48.575	58	28.660	8.519	19.000	48.700	53
	1	53.611	16.411	30.650	85.050	54	46.716	14.259	26.650	78.050	67
	2	75.127	15.041	50.250	97.875	126	71.057	13.957	43.000	92.400	105
IAS02	0	36.600	15.420	19.000	80.325	90	36.158	14.486	19.000	74.000	101
	1	72.693	16.107	41.450	97.275	150	67.256	15.801	37.200	92.000	129
SAS01	0	46.755	19.095	19.000	84.350	147	43.149	17.760	19.675	80.650	148
	1	78.139	13.619	51.775	97.000	72	70.600	13.318	46.600	92.000	65
SAS02	0	40.214	18.239	19.000	84.150	98	38.464	15.948	19.000	78.125	112
	1	70.133	16.017	42.900	96.025	120	66.620	14.559	38.475	92.000	100
ACT01	0	39.066	15.972	19.000	79.750	106	35.352	12.777	19.000	65.650	108
	1	70.392	15.267	41.100	93.425	102	65.241	13.859	41.150	88.850	87
	2	82.677	12.381	56.750	97.250	31	79.415	10.521	60.000	94.000	41
ACT02	0	31.971	11.024	19.000	57.650	68	30.808	9.169	19.000	49.000	73
	1	55.343	14.726	28.450	85.100	70	50.519	14.573	26.000	77.450	52
	2	77.875	14.028	46.125	97.000	96	72.268	14.182	41.800	92.000	97
ACT03	0	31.890	10.303	19.000	51.800	73	33.011	11.086	19.000	60.000	92
	1	61.190	15.087	36.800	88.200	79	60.358	14.671	36.000	87.000	81
	2	81.296	11.871	55.325	99.025	54	78.412	11.779	52.650	96.000	34

Table B.6 Descriptive age-related statistics for standard skeletal markers. Pooled sexes.

		Left					Right				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
CLV01	0	22.712	2.995	19.000	28.550	59	22.661	2.963	19.000	28.550	59
	1	42.543	11.297	26.000	69.925	164	43.048	11.814	26.000	73.300	168
	2	73.992	13.278	50.000	96.000	241	74.444	13.578	50.000	96.000	241
CLV02	0	36.899	15.117	19.000	74.325	188	37.183	14.959	19.000	76.125	186
	1	69.778	16.570	38.000	96.000	230	69.723	16.819	38.000	96.000	249
RB101	0	24.212	5.104	19.000	36.875	66	24.303	5.561	19.000	38.750	66
	1	53.118	17.497	27.000	88.000	246	52.373	16.833	26.875	88.000	236
	2	78.768	11.805	55.000	97.000	112	78.741	11.692	56.350	97.325	108
RB102	0	34.561	12.249	19.000	63.700	173	34.465	12.185	19.000	63.725	172
	1	69.647	15.664	39.000	96.000	235	69.565	15.401	41.850	95.000	239
PSY01	0	24.662	5.212	19.000	38.000	65	24.484	5.017	19.000	37.475	62
	1	56.123	17.891	28.275	89.725	252	55.398	17.458	27.325	89.000	254
	2	76.410	13.285	52.150	95.950	83	77.382	13.444	50.400	95.800	89
PSY02	0	28.396	9.133	19.000	52.000	91	27.739	7.798	19.000	46.475	88
	1	58.393	17.580	29.000	90.925	244	57.808	17.239	29.000	89.775	250
	2	77.896	11.942	55.000	95.350	67	79.083	11.669	55.775	95.225	72
PSY03	0	24.918	7.455	19.000	43.500	61	25.220	7.879	19.000	47.200	59
	1	50.697	16.857	27.000	87.000	178	50.168	17.252	26.450	87.550	179
	2	71.445	15.064	42.400	96.000	155	71.457	15.148	39.100	94.975	162
IAS01	0	29.632	9.282	19.000	49.000	114	29.234	8.932	19.000	49.000	111
	1	51.157	15.946	27.150	81.000	127	49.793	15.577	28.000	81.000	121
	2	73.165	14.864	43.000	96.575	218	73.277	14.668	44.500	97.000	231
IAS02	0	36.290	14.474	19.000	74.600	193	36.366	14.895	19.000	77.250	191
	1	70.007	16.287	38.875	96.000	276	70.179	16.168	39.000	96.050	279
SAS01	0	45.304	18.872	19.000	86.250	296	44.946	18.495	19.000	83.650	295
	1	74.104	13.836	49.650	96.000	134	74.562	13.949	49.800	96.000	137
SAS02	0	38.592	16.403	19.000	81.875	206	39.281	17.036	19.000	82.775	210
	1	68.633	15.303	39.000	94.575	218	68.536	15.437	39.000	95.525	220
ACT01	0	37.304	14.344	19.000	73.200	217	37.192	14.534	19.000	76.025	214
	1	67.922	14.879	40.450	93.650	179	68.021	14.822	40.700	92.000	189
	2	80.448	10.895	60.000	96.000	87	80.819	11.392	59.550	96.225	72
ACT02	0	31.556	10.275	19.000	50.475	142	31.369	10.086	19.000	50.500	141
	1	53.782	14.881	26.300	83.100	133	53.287	14.796	26.025	84.000	122
	2	75.838	13.565	46.750	96.000	191	75.057	14.347	43.000	96.000	193
ACT03	0	32.618	10.767	19.000	60.000	165	32.515	10.729	19.000	60.000	165
	1	60.601	15.164	33.000	89.000	158	60.769	14.837	35.925	88.025	160
	2	79.681	12.082	53.000	96.000	94	80.182	11.852	53.350	96.825	88

Table B.7 Descriptive age-related statistics for cranial and palatine sutures.

	Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
Pooled	CRS01	0	41.836	20.943	19.000	90.025	134
		1	62.409	20.469	24.050	94.000	323
	CRS02L	0	52.493	22.492	19.800	92.200	353
		1	68.198	19.049	29.500	94.000	101
	CRS02R	0	52.606	22.545	19.800	92.200	353
		1	67.600	19.139	29.475	94.000	100
	CRS03	0	47.855	22.310	19.000	90.300	235
		1	66.567	19.626	28.150	96.000	247
	CRS04L	0	42.057	21.076	19.000	88.000	174
		1	65.981	19.102	28.675	96.000	308
	CRS04R	0	41.263	20.667	19.000	87.750	171
		1	66.370	18.955	29.000	96.000	311
	CRS05	0	45.188	22.815	19.000	92.000	181
		1	64.594	19.457	28.000	96.000	283
	CRS06L	0	49.688	22.486	19.000	92.000	266
		1	65.930	19.922	27.950	96.000	200
	CRS06R	0	50.034	22.805	19.000	92.000	266
		1	65.768	19.731	28.000	96.000	198
Male	CRS01	0	29.469	14.830	19.000	70.175	32
		1	58.910	20.479	24.000	92.000	201
	CRS02L	0	48.877	21.680	19.825	92.000	154
		1	66.270	18.817	30.825	92.175	74
	CRS02R	0	48.883	21.692	19.825	92.000	154
		1	65.959	18.755	30.800	92.200	73
	CRS03	0	46.330	22.338	19.000	87.000	106
		1	63.030	19.188	28.275	92.725	132
	CRS04L	0	37.333	18.008	19.000	87.000	78
		1	64.204	18.476	29.000	92.000	162
	CRS04R	0	36.853	18.066	19.000	87.000	75
		1	63.933	18.475	29.000	92.000	165
	CRS05	0	41.813	22.136	19.000	88.600	75
		1	61.772	19.123	27.000	92.075	158
	CRS06L	0	48.229	21.766	19.250	88.000	131
		1	63.693	19.541	29.500	92.500	101
	CRS06R	0	48.362	21.818	19.225	88.000	130
		1	63.673	19.544	29.500	92.500	101
Female	CRS01	0	45.716	21.128	19.000	91.475	102
		1	68.172	19.181	28.025	95.000	122
	CRS02L	0	55.291	22.761	19.950	95.050	199
		1	73.481	19.025	21.600	94.350	27
	CRS02R	0	55.487	22.823	19.950	95.050	199
		1	72.037	19.817	21.600	94.350	27
	CRS03	0	49.109	22.296	19.000	91.600	129
		1	70.626	19.415	28.850	98.150	115
	CRS04L	0	45.896	22.647	19.000	91.625	96
		1	67.952	19.649	29.250	98.000	146
	CRS04R	0	44.708	21.972	19.000	91.625	96
		1	69.123	19.175	30.625	98.000	146
	CRS05	0	47.575	23.089	19.000	93.750	106
		1	68.160	19.365	30.300	97.900	125
	CRS06L	0	51.104	23.156	19.000	94.300	135
		1	68.212	20.147	26.000	98.100	99
	CRS06R	0	51.632	23.679	19.000	94.250	136
		1	67.948	19.790	26.800	98.200	97

Table B.8 Descriptive sex-specific age-related statistics vertebrae traits.

		Female					Male				
Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
C3IS	0	38.447	15.259	19.000	77.850	103	36.477	13.591	19.000	67.100	107
	1	68.662	14.072	44.825	92.700	74	65.014	14.297	38.000	90.600	73
	2	84.457	10.538	60.250	99.875	46	78.047	10.576	60.000	93.900	43
C4SS	0	37.010	13.903	19.000	68.125	96	36.135	13.330	19.000	65.850	104
	1	66.600	14.294	38.950	92.000	80	65.139	15.425	37.750	90.100	79
	2	86.286	8.715	69.600	99.800	49	77.044	11.909	55.100	95.800	45
C4IS	0	35.728	12.708	19.000	64.725	92	35.868	12.448	19.000	64.375	106
	1	67.273	13.997	44.900	94.400	77	63.466	13.498	38.000	90.400	73
	2	83.143	12.615	51.125	98.625	56	80.167	10.680	56.400	95.650	48
C5SS	0	34.565	12.182	19.000	63.800	85	35.904	13.112	19.000	65.850	104
	1	65.195	15.256	38.800	97.100	77	61.492	13.670	38.000	87.800	65
	2	82.385	12.441	50.000	98.400	65	78.476	11.150	55.000	94.900	63
C5IS	0	33.694	10.670	19.000	54.800	85	33.711	10.699	19.000	56.600	97
	1	64.750	14.163	44.475	94.625	60	61.393	13.157	38.750	82.250	56
	2	80.012	13.462	50.000	98.900	85	75.538	12.679	49.875	94.050	80
C6SS	0	34.195	11.058	19.000	53.850	87	33.306	10.422	19.000	55.575	98
	1	63.873	13.675	40.750	90.950	55	60.200	13.586	38.000	82.300	55
	2	80.154	13.159	50.000	98.750	91	75.488	12.221	51.100	92.975	82
C6IS	0	35.717	12.555	19.000	64.175	92	33.327	10.815	19.000	55.575	98
	1	63.723	14.893	38.050	97.100	47	60.478	12.956	38.250	80.750	46
	2	79.185	13.180	48.550	97.725	92	73.758	13.313	48.750	92.750	91
C7SS	0	35.034	12.106	19.000	61.125	88	35.070	12.878	19.000	67.575	100
	1	64.933	15.446	37.000	95.150	60	60.500	13.293	38.000	80.750	46
	2	79.301	13.397	48.100	97.950	83	74.888	13.468	49.000	93.800	89
L1IS	0	38.113	13.759	19.000	65.750	106	34.602	12.256	19.000	62.875	98
	1	66.694	15.023	38.050	92.375	62	61.859	13.366	38.750	89.250	71
	2	82.684	10.163	61.400	97.600	57	78.526	11.296	54.000	95.200	57
L2SS	0	36.845	13.122	19.000	63.800	97	34.091	11.736	19.000	60.000	99
	1	62.967	15.256	34.900	89.525	60	62.238	13.126	39.550	87.900	63
	2	81.522	10.737	61.700	97.300	69	76.712	11.968	49.625	94.750	66
L2IS	0	37.089	12.872	19.000	63.500	101	32.968	10.296	19.000	56.000	95
	1	63.877	14.159	37.800	91.200	57	61.333	12.700	39.625	90.125	66
	2	81.939	10.877	57.625	97.375	66	77.597	11.176	53.250	94.700	67
L3SS	0	34.048	11.586	19.000	63.450	83	32.533	10.159	19.000	56.000	92
	1	60.388	14.109	35.600	90.700	67	58.463	12.687	38.000	83.800	67
	2	79.947	12.460	50.000	97.150	75	77.270	10.980	54.650	94.350	74
L3IS	0	35.415	11.854	19.000	61.075	94	33.371	10.592	19.000	56.000	97
	1	63.766	13.151	41.300	92.125	64	60.377	12.832	38.000	84.200	69
	2	81.328	10.897	56.300	97.350	67	78.197	10.876	55.000	94.750	66
L4SS	0	32.885	10.322	19.000	51.075	78	32.033	9.354	19.000	49.000	91
	1	57.683	12.863	33.000	80.900	63	58.655	12.267	38.425	81.000	58
	2	78.953	12.665	50.625	97.000	86	75.470	12.605	52.150	93.950	83
L4IS	0	35.817	13.023	19.000	63.500	93	33.459	10.684	19.000	58.300	98
	1	61.660	13.683	37.000	89.700	53	59.915	12.218	38.450	81.000	59
	2	79.325	13.210	50.150	97.000	83	76.329	12.792	49.500	94.250	76
L5SS	0	32.938	10.475	19.000	51.050	80	33.085	10.639	19.000	58.700	94
	1	56.882	13.046	33.000	79.000	51	59.018	12.118	38.375	81.000	56
	2	77.284	13.564	50.000	97.000	95	74.556	13.276	46.000	93.000	81
S1SS	0	33.810	11.220	19.000	53.925	84	33.221	10.039	19.000	53.600	95
	1	62.508	15.371	38.000	91.250	63	62.704	14.295	36.500	89.250	71
	2	77.568	13.477	49.125	98.350	74	74.271	12.825	49.450	92.550	59
S1S2F	0	25.692	8.527	19.000	48.125	26	27.065	7.172	19.000	42.875	46
	1	61.986	21.354	25.000	95.800	209	60.635	18.797	26.000	92.000	189

Table B.9 Descriptive sex-specific age-related statistics vertebrae traits.

Trait	Stage	mean(y)	sd(y)	$Q_{0.025}$(y)	$Q_{0.975}$ (y)	n
C3IS	0	37.443	14.432	19.000	71.775	210
	1	66.850	14.254	39.300	92.350	147
	2	81.360	10.980	59.200	98.800	89
C4SS	0	36.555	13.581	19.000	67.075	200
	1	65.874	14.837	37.950	92.000	159
	2	81.862	11.307	57.300	98.675	94
C4IS	0	35.803	12.538	19.000	65.000	198
	1	65.420	13.842	39.450	92.275	150
	2	81.769	11.800	51.725	97.425	104
C5SS	0	35.302	12.686	19.000	65.000	189
	1	63.500	14.618	38.000	92.000	142
	2	80.461	11.939	50.525	97.000	128
C5IS	0	33.703	10.656	19.000	56.475	182
	1	63.129	13.731	39.750	92.000	116
	2	77.842	13.240	50.000	97.000	165
C6SS	0	33.724	10.706	19.000	55.400	185
	1	62.036	13.693	38.000	89.825	110
	2	77.942	12.900	50.000	97.000	173
C6IS	0	34.484	11.720	19.000	62.550	190
	1	62.118	13.988	38.000	91.400	93
	2	76.486	13.487	48.000	96.450	183
C7SS	0	35.053	12.489	19.000	66.975	188
	1	63.009	14.651	37.000	91.375	106
	2	77.017	13.576	48.000	96.725	172
L1IS	0	36.426	13.145	19.000	65.000	204
	1	64.113	14.314	38.300	90.000	133
	2	80.605	10.899	59.650	96.175	114
L2SS	0	35.454	12.486	19.000	62.375	196
	1	62.593	14.151	38.000	89.000	123
	2	79.170	11.567	55.700	96.650	135
L2IS	0	35.092	11.846	19.000	60.250	196
	1	62.512	13.401	39.000	91.900	123
	2	79.752	11.201	52.900	96.700	133
L3SS	0	33.251	10.854	19.000	58.600	175
	1	59.425	13.401	37.325	89.675	134
	2	78.617	11.785	51.400	96.300	149
L3IS	0	34.377	11.248	19.000	57.000	191
	1	62.008	13.048	38.300	88.400	133
	2	79.774	10.958	55.000	96.000	133
L4SS	0	32.426	9.792	19.000	50.800	169
	1	58.149	12.538	33.000	81.000	121
	2	77.243	12.718	50.400	96.000	169
L4IS	0	34.607	11.907	19.000	60.250	191
	1	60.741	12.903	37.000	84.350	112
	2	77.893	13.058	49.800	96.050	159
L5SS	0	33.017	10.534	19.000	55.025	174
	1	58.000	12.555	33.000	81.000	107
	2	76.028	13.463	47.125	96.000	176
S1SS	0	33.497	10.584	19.000	54.550	179
	1	62.612	14.755	38.000	89.675	134
	2	76.105	13.244	49.300	96.700	133
S1S2F	0	26.569	7.658	19.000	47.225	72
	1	61.344	20.166	26.000	95.000	398

Age-related correlation analysis

Table B.10 Age-related correlation analysis for bilateral traits. Female individuals. Left side.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
SC01	0.801	0.644	156.433	0.000	243	0.748	0.539	
HM01	0.662	0.440	106.532	0.000	242	0.405	0.488	
HM02	0.760	0.579	139.019	0.000	240	0.533	0.631	
HM03	0.659	0.433	103.005	0.000	238	0.326	0.571	
HM04	0.446	0.199	47.187	0.000	237	0.048	0.763	
HM05	0.688	0.477	105.964	0.000	222	0.280	0.747	
HM06	0.769	0.592	134.327	0.000	227	0.463	0.737	
UL01	0.616	0.380	90.479	0.000	238	0.229	0.644	
UL02	0.551	0.304	68.479	0.000	225	0.161	0.607	
RD01	0.402	0.161	37.271	0.000	232	0.038	0.641	
RD02	0.744	0.559	131.854	0.000	236	0.615	0.498	
OC01	0.740	0.543	119.454	0.000	220	0.452	0.648	
OC02	0.738	0.537	122.532	0.000	228	0.708	0.380	
OC03	0.805	0.650	157.251	0.000	242	0.726	0.570	
FM01	0.470	0.223	53.580	0.000	240	0.102	0.532	
FM02	0.464	0.216	49.284	0.000	228	0.070	0.674	
FM03	0.768	0.578	133.027	0.000	230	0.497	0.671	
FM04	0.767	0.582	131.527	0.000	226	0.504	0.667	
FM05	0.649	0.424	100.151	0.000	236	0.313	0.586	
TB01	0.550	0.304	72.861	0.000	240	0.172	0.562	
PT01	0.631	0.395	88.961	0.000	225	0.231	0.652	
PT02	0.791	0.611	125.906	0.000	206	0.612	0.610	
CLN01	0.769	0.589	124.944	0.000	212	0.595	0.583	
CLV01	0.841	0.712	163.679	0.000	230	0.996	0.599	0.613
CLV02	0.753	0.554	114.137	0.000	206	0.640	0.473	
RB101	0.748	0.568	119.932	0.000	211	0.965	0.001	0.802
RB102	0.774	0.593	125.112	0.000	211	0.696	0.496	
PSY01	0.678	0.472	96.293	0.000	204	0.971	0.004	0.738
PSY02	0.703	0.494	100.862	0.000	204	0.907	0.024	0.792
PSY03	0.730	0.542	109.003	0.000	201	0.937	0.105	0.521
IAS01	0.786	0.634	149.525	0.000	236	0.890	0.077	0.550
IAS02	0.739	0.554	132.872	0.000	240	0.673	0.434	
SAS01	0.624	0.390	84.199	0.000	216	0.192	0.713	
SAS02	0.666	0.439	93.603	0.000	213	0.460	0.417	
ACT01	0.756	0.585	140.516	0.000	240	0.616	0.417	0.784
ACT02	0.828	0.687	161.416	0.000	235	0.838	0.011	0.742
ACT03	0.824	0.681	137.628	0.000	202	0.836	0.094	0.794

Table B.11 Age-related correlation analysis for bilateral traits. Female individuals. Right side.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
SC01	0.774	0.602	147.455	0.000	245	0.713	0.494	
HM01	0.605	0.366	87.196	0.000	238	0.298	0.475	
HM02	0.771	0.594	138.323	0.000	233	0.592	0.596	
HM03	0.703	0.493	114.269	0.000	232	0.419	0.578	
HM04	0.458	0.210	49.385	0.000	235	0.054	0.742	
HM05	0.704	0.506	110.222	0.000	218	0.306	0.760	
HM06	0.762	0.574	125.623	0.000	219	0.470	0.689	
UL01	0.679	0.459	109.784	0.000	239	0.334	0.633	
UL02	0.567	0.321	71.892	0.000	224	0.172	0.618	
RD01	0.396	0.157	36.381	0.000	232	0.038	0.613	
RD02	0.732	0.541	127.140	0.000	235	0.562	0.517	
OC01	0.737	0.538	118.903	0.000	221	0.462	0.626	
OC02	0.749	0.554	126.349	0.000	228	0.723	0.396	
OC03	0.784	0.618	150.054	0.000	243	0.654	0.577	
FM01	0.494	0.245	58.465	0.000	239	0.118	0.543	
FM02	0.466	0.218	50.941	0.000	234	0.066	0.695	
FM03	0.767	0.580	131.589	0.000	227	0.512	0.653	
FM04	0.749	0.556	127.990	0.000	230	0.480	0.646	
FM05	0.657	0.432	101.623	0.000	235	0.326	0.588	
TB01	0.538	0.290	69.321	0.000	239	0.154	0.577	
PT01	0.645	0.411	91.627	0.000	223	0.269	0.627	
PT02	0.766	0.578	120.195	0.000	208	0.539	0.619	
CLN01	0.766	0.578	121.428	0.000	210	0.586	0.570	
CLV01	0.831	0.696	164.263	0.000	236	0.996	0.573	0.597
CLV02	0.757	0.569	124.003	0.000	218	0.698	0.452	
RB101	0.766	0.593	122.670	0.000	207	0.971	0.006	0.794
RB102	0.765	0.575	121.416	0.000	211	0.689	0.473	
PSY01	0.693	0.495	101.502	0.000	205	0.975	0.000	0.723
PSY02	0.718	0.523	106.608	0.000	204	0.940	0.013	0.788
PSY03	0.729	0.540	108.490	0.000	201	0.937	0.142	0.506
IAS01	0.784	0.638	151.953	0.000	238	0.910	0.105	0.531
IAS02	0.733	0.551	132.135	0.000	240	0.684	0.415	
SAS01	0.647	0.418	91.440	0.000	219	0.227	0.708	
SAS02	0.659	0.435	94.765	0.000	218	0.452	0.416	
ACT01	0.750	0.569	136.014	0.000	239	0.589	0.396	0.803
ACT02	0.816	0.669	156.542	0.000	234	0.848	0.027	0.674
ACT03	0.837	0.706	145.539	0.000	206	0.848	0.104	0.821

Table B.12 Age-related correlation analysis for bilateral traits. Male individuals. Left side.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
SC01	0.769	0.600	143.922	0.000	240	0.677	0.526	
HM01	0.757	0.578	136.994	0.000	237	0.530	0.627	
HM02	0.666	0.448	105.232	0.000	235	0.460	0.435	
HM03	0.513	0.264	59.596	0.000	226	0.148	0.471	
HM04	0.453	0.211	50.311	0.000	239	0.060	0.661	
HM05	0.637	0.416	89.761	0.000	216	0.199	0.768	
HM06	0.691	0.483	102.793	0.000	213	0.277	0.778	
UL01	0.528	0.285	67.564	0.000	237	0.126	0.620	
UL02	0.558	0.312	69.873	0.000	224	0.222	0.442	
RD01	0.390	0.155	36.321	0.000	234	0.033	0.645	
RD02	0.720	0.520	124.865	0.000	240	0.632	0.424	
OC01	0.683	0.459	97.802	0.000	213	0.459	0.459	
OC02	0.695	0.485	115.466	0.000	238	0.549	0.426	
OC03	0.766	0.590	143.453	0.000	243	0.600	0.581	
FM01	0.680	0.472	113.199	0.000	240	0.257	0.778	
FM02	0.550	0.303	68.868	0.000	227	0.117	0.712	
FM03	0.753	0.576	125.539	0.000	218	0.483	0.666	
FM04	0.762	0.581	133.688	0.000	230	0.652	0.519	
FM05	0.667	0.448	107.511	0.000	240	0.307	0.639	
TB01	0.608	0.374	87.428	0.000	234	0.205	0.641	
PT01	0.622	0.388	84.101	0.000	217	0.222	0.649	
PT02	0.632	0.408	83.592	0.000	205	0.246	0.624	
CLN01	0.731	0.540	124.756	0.000	231	0.476	0.608	
CLV01	0.872	0.754	176.423	0.000	234	0.990	0.609	0.669
CLV02	0.691	0.476	100.923	0.000	212	0.540	0.417	
RB101	0.790	0.614	130.848	0.000	213	0.988	0.057	0.804
RB102	0.787	0.609	119.894	0.000	197	0.785	0.463	
PSY01	0.734	0.553	108.340	0.000	196	0.974	0.001	0.718
PSY02	0.751	0.568	112.552	0.000	198	0.883	0.062	0.803
PSY03	0.706	0.519	100.190	0.000	193	0.950	0.024	0.584
IAS01	0.787	0.613	136.725	0.000	223	0.899	0.095	0.587
IAS02	0.722	0.514	117.712	0.000	229	0.628	0.415	
SAS01	0.599	0.357	76.321	0.000	214	0.180	0.655	
SAS02	0.718	0.510	107.640	0.000	211	0.479	0.543	
ACT01	0.818	0.676	164.241	0.000	243	0.674	0.334	0.867
ACT02	0.822	0.676	156.211	0.000	231	0.868	0.067	0.655
ACT03	0.818	0.671	144.160	0.000	215	0.753	0.250	0.850

Table B.13 Age-related correlation analysis for bilateral traits. Male individuals. Right side.

Trait	ρ	η^2	χ^2	p-value	n	Stage		
						0	1	2
SC01	0.768	0.594	144.367	0.000	243	0.648	0.540	
HM01	0.772	0.601	141.772	0.000	236	0.577	0.625	
HM02	0.679	0.461	108.839	0.000	236	0.543	0.386	
HM03	0.582	0.342	75.895	0.000	222	0.251	0.472	
HM04	0.471	0.227	53.004	0.000	234	0.081	0.609	
HM05	0.605	0.370	79.515	0.000	215	0.182	0.708	
HM06	0.732	0.537	113.944	0.000	212	0.368	0.752	
UL01	0.537	0.292	69.138	0.000	237	0.134	0.624	
UL02	0.550	0.305	66.862	0.000	219	0.239	0.387	
RD01	0.441	0.203	47.292	0.000	233	0.045	0.734	
RD02	0.684	0.470	109.603	0.000	233	0.580	0.375	
OC01	0.735	0.542	113.371	0.000	209	0.562	0.526	
OC02	0.730	0.530	124.631	0.000	235	0.587	0.478	
OC03	0.757	0.577	139.118	0.000	241	0.538	0.619	
FM01	0.669	0.458	108.915	0.000	238	0.252	0.760	
FM02	0.512	0.266	59.322	0.000	223	0.096	0.663	
FM03	0.724	0.532	117.525	0.000	221	0.426	0.639	
FM04	0.754	0.566	130.239	0.000	230	0.638	0.502	
FM05	0.694	0.488	116.075	0.000	238	0.322	0.700	
TB01	0.618	0.386	90.333	0.000	234	0.211	0.658	
PT01	0.605	0.368	79.496	0.000	216	0.204	0.642	
PT02	0.579	0.332	68.110	0.000	205	0.176	0.603	
CLN01	0.742	0.557	125.851	0.000	226	0.482	0.635	
CLV01	0.870	0.748	173.546	0.000	232	0.990	0.575	0.672
CLV02	0.667	0.438	95.118	0.000	217	0.516	0.368	
RB101	0.797	0.626	127.144	0.000	203	0.966	0.050	0.831
RB102	0.802	0.637	127.308	0.000	200	0.803	0.488	
PSY01	0.732	0.549	109.836	0.000	200	0.975	0.000	0.737
PSY02	0.761	0.583	120.046	0.000	206	0.914	0.040	0.838
PSY03	0.695	0.496	98.633	0.000	199	0.936	0.034	0.579
IAS01	0.811	0.645	145.188	0.000	225	0.899	0.201	0.606
IAS02	0.719	0.509	116.961	0.000	230	0.594	0.429	
SAS01	0.609	0.371	79.008	0.000	213	0.183	0.676	
SAS02	0.686	0.460	97.465	0.000	212	0.412	0.513	
ACT01	0.808	0.661	156.091	0.000	236	0.683	0.399	0.857
ACT02	0.818	0.666	147.885	0.000	222	0.862	0.042	0.638
ACT03	0.809	0.654	135.430	0.000	207	0.731	0.284	0.846

Table B.14 Age-related correlation analysis for bilateral traits. Left side. Pooled sexes.

Trait	ρ	η^2	χ^2	p-value	n	Stage		
						0	1	2
SC01	0.786	0.621	299.937	0.000	483	0.715	0.530	
HM01	0.711	0.506	242.225	0.000	479	0.461	0.559	
HM02	0.711	0.508	241.233	0.000	475	0.497	0.519	
HM03	0.592	0.351	162.690	0.000	464	0.232	0.525	
HM04	0.447	0.201	95.583	0.000	476	0.053	0.699	
HM05	0.664	0.448	196.437	0.000	438	0.239	0.755	
HM06	0.734	0.542	238.675	0.000	440	0.371	0.754	
UL01	0.579	0.337	160.169	0.000	475	0.177	0.635	
UL02	0.549	0.301	135.090	0.000	449	0.184	0.506	
RD01	0.398	0.159	73.872	0.000	466	0.036	0.642	
RD02	0.731	0.537	255.505	0.000	476	0.622	0.456	
OC01	0.703	0.490	212.101	0.000	433	0.449	0.532	
OC02	0.719	0.511	238.305	0.000	466	0.628	0.405	
OC03	0.788	0.621	301.273	0.000	485	0.667	0.575	
FM01	0.574	0.330	158.438	0.000	480	0.161	0.664	
FM02	0.504	0.254	115.543	0.000	455	0.089	0.692	
FM03	0.761	0.577	258.586	0.000	448	0.491	0.667	
FM04	0.754	0.567	258.688	0.000	456	0.560	0.574	
FM05	0.660	0.436	207.728	0.000	476	0.311	0.612	
TB01	0.581	0.337	159.556	0.000	474	0.187	0.602	
PT01	0.624	0.388	171.659	0.000	442	0.225	0.642	
PT02	0.714	0.510	209.626	0.000	411	0.408	0.621	
CLN01	0.750	0.565	250.121	0.000	443	0.534	0.597	
CLV01	0.856	0.731	339.234	0.000	464	0.992	0.601	0.639
CLV02	0.721	0.514	214.975	0.000	418	0.590	0.444	
RB101	0.763	0.579	245.401	0.000	424	0.974	0.017	0.798
RB102	0.777	0.596	243.084	0.000	408	0.732	0.475	
PSY01	0.701	0.504	201.663	0.000	400	0.972	0.003	0.720
PSY02	0.725	0.527	211.667	0.000	402	0.895	0.039	0.791
PSY03	0.720	0.531	209.330	0.000	394	0.943	0.058	0.549
IAS01	0.786	0.620	284.625	0.000	459	0.893	0.094	0.565
IAS02	0.732	0.533	249.933	0.000	469	0.654	0.421	
SAS01	0.607	0.370	158.909	0.000	430	0.185	0.674	
SAS02	0.694	0.474	201.072	0.000	424	0.471	0.478	
ACT01	0.782	0.620	299.271	0.000	483	0.641	0.375	0.831
ACT02	0.818	0.670	312.293	0.000	466	0.851	0.023	0.681
ACT03	0.823	0.677	282.515	0.000	417	0.796	0.167	0.822

Table B.15 Age-related correlation analysis for bilateral traits. Right side. Pooled sexes.

Trait	ρ	η^2	χ^2	p-value	n	Stage		
						0	1	2
SC01	0.772	0.597	291.442	0.000	488	0.682	0.514	
HM01	0.687	0.472	223.630	0.000	474	0.410	0.553	
HM02	0.721	0.519	243.468	0.000	469	0.566	0.474	
HM03	0.646	0.418	189.732	0.000	454	0.333	0.524	
HM04	0.458	0.211	98.870	0.000	469	0.065	0.654	
HM05	0.658	0.438	189.810	0.000	433	0.241	0.730	
HM06	0.747	0.557	240.091	0.000	431	0.420	0.718	
UL01	0.616	0.380	180.950	0.000	476	0.226	0.630	
UL02	0.551	0.303	134.359	0.000	443	0.196	0.474	
RD01	0.420	0.179	83.074	0.000	465	0.042	0.674	
RD02	0.707	0.502	234.942	0.000	468	0.568	0.436	
OC01	0.730	0.530	227.769	0.000	430	0.497	0.563	
OC02	0.744	0.543	251.609	0.000	463	0.656	0.439	
OC03	0.774	0.599	290.138	0.000	484	0.601	0.598	
FM01	0.581	0.338	161.419	0.000	477	0.172	0.658	
FM02	0.483	0.235	107.440	0.000	457	0.077	0.670	
FM03	0.747	0.558	249.884	0.000	448	0.473	0.646	
FM04	0.740	0.546	251.094	0.000	460	0.538	0.554	
FM05	0.678	0.461	218.258	0.000	473	0.326	0.646	
TB01	0.579	0.333	157.714	0.000	473	0.178	0.619	
PT01	0.628	0.392	171.984	0.000	439	0.239	0.631	
PT02	0.680	0.459	189.666	0.000	413	0.336	0.614	
CLN01	0.755	0.568	247.703	0.000	436	0.534	0.604	
CLV01	0.851	0.721	337.241	0.000	468	0.993	0.571	0.633
CLV02	0.713	0.503	218.646	0.000	435	0.609	0.407	
RB101	0.777	0.600	245.871	0.000	410	0.968	0.020	0.809
RB102	0.782	0.601	246.972	0.000	411	0.738	0.477	
PSY01	0.712	0.520	210.436	0.000	405	0.975	0.000	0.728
PSY02	0.740	0.551	225.873	0.000	410	0.926	0.024	0.811
PSY03	0.713	0.518	207.068	0.000	400	0.936	0.076	0.540
IAS01	0.798	0.636	294.683	0.000	463	0.904	0.161	0.565
IAS02	0.729	0.530	249.233	0.000	470	0.646	0.420	
SAS01	0.624	0.392	169.545	0.000	432	0.206	0.679	
SAS02	0.677	0.449	193.072	0.000	430	0.437	0.463	
ACT01	0.775	0.607	288.208	0.000	475	0.629	0.395	0.827
ACT02	0.813	0.661	301.429	0.000	456	0.854	0.027	0.646
ACT03	0.828	0.685	282.916	0.000	413	0.796	0.189	0.836

Table B.16 Age-related correlation analysis for cranial and vertebrae traits. Female individuals.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
CRS01	0.487	0.238	53.353	0.000	224	0.253	0.223	
CRS02L	0.257	0.066	14.833	0.000	226	0.009	0.424	
CRS02R	0.232	0.054	12.278	0.000	226	0.007	0.360	
CRS03	0.456	0.209	50.974	0.000	244	0.173	0.257	
CRS04L	0.453	0.212	51.312	0.000	242	0.259	0.166	
CRS04R	0.500	0.258	62.488	0.000	242	0.312	0.204	
CRS05	0.437	0.192	44.285	0.000	231	0.190	0.193	
CRS06L	0.359	0.130	30.480	0.000	234	0.090	0.195	
CRS06R	0.339	0.117	27.359	0.000	233	0.077	0.190	
C3IS	0.803	0.649	144.795	0.000	223	0.623	0.369	0.866
C4SS	0.834	0.696	156.630	0.000	225	0.702	0.256	0.913
C4IS	0.829	0.696	156.523	0.000	225	0.762	0.293	0.798
C5SS	0.824	0.690	156.663	0.000	227	0.798	0.157	0.787
C5IS	0.833	0.720	165.487	0.000	230	0.850	0.147	0.713
C6SS	0.839	0.725	168.983	0.000	233	0.837	0.108	0.720
C6IS	0.820	0.687	158.590	0.000	231	0.773	0.103	0.709
C7SS	0.809	0.676	156.252	0.000	231	0.795	0.142	0.705
L1IS	0.817	0.673	151.468	0.000	225	0.662	0.285	0.864
L2SS	0.827	0.685	154.888	0.000	226	0.713	0.118	0.836
L2IS	0.836	0.702	157.356	0.000	224	0.710	0.188	0.841
L3SS	0.838	0.702	157.969	0.000	225	0.802	0.050	0.772
L3IS	0.853	0.731	164.453	0.000	225	0.773	0.204	0.833
L4SS	0.854	0.730	165.736	0.000	227	0.849	0.001	0.749
L4IS	0.821	0.680	155.675	0.000	229	0.738	0.084	0.733
L5SS	0.843	0.712	161.011	0.000	226	0.842	0.000	0.694
S1SS	0.811	0.671	148.376	0.000	221	0.807	0.129	0.710
S1S2F	0.487	0.240	56.306	0.000	235	0.937	0.034	

Table B.17 Age-related correlation analysis for cranial and vertebrae traits. Male individuals.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
CRS01	0.465	0.209	48.640	0.000	233	0.752	0.038	
CRS02L	0.369	0.134	30.541	0.000	228	0.064	0.283	
CRS02R	0.362	0.129	29.330	0.000	227	0.061	0.279	
CRS03	0.373	0.140	33.346	0.000	238	0.148	0.131	
CRS04L	0.569	0.322	77.354	0.000	240	0.507	0.184	
CRS04R	0.567	0.321	76.943	0.000	240	0.518	0.174	
CRS05	0.423	0.178	41.432	0.000	233	0.275	0.102	
CRS06L	0.353	0.120	27.898	0.000	232	0.088	0.168	
CRS06R	0.348	0.118	27.259	0.000	231	0.087	0.164	
C3IS	0.789	0.631	140.624	0.000	223	0.622	0.383	0.843
C4SS	0.772	0.608	138.648	0.000	228	0.651	0.335	0.789
C4IS	0.821	0.681	154.634	0.000	227	0.684	0.328	0.859
C5SS	0.814	0.668	155.079	0.000	232	0.673	0.204	0.823
C5IS	0.832	0.707	164.633	0.000	233	0.796	0.207	0.732
C6SS	0.838	0.715	168.100	0.000	235	0.804	0.160	0.752
C6IS	0.824	0.695	163.267	0.000	235	0.792	0.189	0.684
C7SS	0.802	0.654	153.741	0.000	235	0.710	0.143	0.685
L1IS	0.824	0.688	155.536	0.000	226	0.722	0.248	0.825
L2SS	0.828	0.698	159.237	0.000	228	0.748	0.276	0.782
L2IS	0.855	0.743	169.353	0.000	228	0.813	0.238	0.815
L3SS	0.859	0.744	173.380	0.000	233	0.821	0.103	0.817
L3IS	0.854	0.736	170.777	0.000	232	0.796	0.193	0.832
L4SS	0.854	0.739	171.418	0.000	232	0.851	0.117	0.742
L4IS	0.843	0.718	167.187	0.000	233	0.791	0.185	0.753
L5SS	0.832	0.701	161.980	0.000	231	0.795	0.153	0.710
S1SS	0.815	0.679	152.809	0.000	225	0.802	0.306	0.731
S1S2F	0.627	0.378	88.718	0.000	235	0.935	0.109	

Table B.18 Age-related correlation analysis for cranial and vertebrae traits. Pooled sexes.

						Stage		
Trait	ρ	η^2	χ^2	p-value	n	0	1	2
CRS01	0.415	0.172	78.510	0.000	457	0.327	0.080	
CRS02L	0.288	0.083	37.625	0.000	454	0.024	0.293	
CRS02R	0.275	0.075	34.104	0.000	453	0.021	0.273	
CRS03	0.406	0.166	80.204	0.000	482	0.156	0.178	
CRS04L	0.498	0.252	121.468	0.000	482	0.346	0.170	
CRS04R	0.520	0.274	132.216	0.000	482	0.382	0.181	
CRS05	0.413	0.172	79.702	0.000	464	0.213	0.132	
CRS06L	0.352	0.124	57.733	0.000	466	0.088	0.179	
CRS06R	0.339	0.116	53.741	0.000	464	0.080	0.174	
C3IS	0.795	0.637	283.956	0.000	446	0.622	0.373	0.845
C4SS	0.802	0.649	293.865	0.000	453	0.679	0.298	0.838
C4IS	0.827	0.688	311.124	0.000	452	0.727	0.308	0.826
C5SS	0.821	0.678	311.263	0.000	459	0.739	0.182	0.801
C5IS	0.833	0.711	329.279	0.000	463	0.825	0.176	0.718
C6SS	0.839	0.718	336.159	0.000	468	0.823	0.132	0.731
C6IS	0.821	0.686	319.529	0.000	466	0.780	0.140	0.688
C7SS	0.805	0.662	308.679	0.000	466	0.755	0.149	0.688
L1IS	0.818	0.674	303.793	0.000	451	0.685	0.256	0.840
L2SS	0.827	0.688	312.414	0.000	454	0.728	0.188	0.804
L2IS	0.844	0.716	323.583	0.000	452	0.752	0.207	0.823
L3SS	0.849	0.722	330.485	0.000	458	0.811	0.073	0.792
L3IS	0.853	0.731	333.932	0.000	457	0.782	0.194	0.830
L4SS	0.854	0.733	336.491	0.000	459	0.850	0.037	0.743
L4IS	0.832	0.697	321.825	0.000	462	0.762	0.127	0.742
L5SS	0.839	0.707	322.872	0.000	457	0.820	0.041	0.702
S1SS	0.815	0.675	301.137	0.000	446	0.805	0.214	0.720
S1S2F	0.561	0.308	144.782	0.000	470	0.937	0.065	

Appendix C

Computational model assessment

Table C.1 Leave-one-out cross-validation for models built on pre-specified skeletal traits sets. Ensembled deep randomized neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.245	0.655	0.953	68.120	51.782	69.796
(m = 9)	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	Median	8.156	0.200	0.960	38.825	33.594	40.881
(m = 16)	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	Median	7.557	0.169	0.960	37.534	29.035	39.599
(m = 16)	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	Median	8.943	0.245	0.963	49.216	17.336	51.768
(m = 2)	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
First rib	Median	9.409	0.275	0.950	48.897	24.356	49.811
(m = 2)	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	Median	10.898	0.370	0.932	51.113	27.029	57.040
(m = 3)	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	Median	8.438	0.220	0.950	44.765	20.350	48.037
(m = 4)	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	Median	8.833	0.229	0.965	43.051	31.541	45.832
(m = 3)	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	Median	6.929	0.147	0.963	33.744	28.816	35.194
(m = 39)	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	Median	6.561	0.145	0.948	34.283	12.952	41.170
(m = 16)	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
Full	Median	5.899	0.118	0.950	30.057	15.558	36.141
(m = 64)	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 C.2 Leave-one-out cross-validation for models built on pre-specified skeletal traits sets. Ensembled deep randomized neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.245	0.655	0.950	68.875	49.061	71.352
(m = 9)	95% CI	14.683	0.616	0.950	66.402	43.918	68.759
		15.751	0.692	0.950	71.000	53.319	73.771
Axial	Median	8.156	0.200	0.950	37.171	33.938	38.585
(m = 16)	95% CI	7.896	0.184	0.950	35.662	32.376	36.846
		8.394	0.213	0.953	39.232	35.124	41.247
Appendicular	Median	7.557	0.169	0.950	35.931	29.751	39.105
(m = 16)	95% CI	7.278	0.155	0.950	34.141	28.298	36.976
		7.823	0.184	0.950	37.521	31.194	41.038
Clavicle	Median	8.943	0.245	0.950	44.947	16.556	50.739
(m = 2)	95% CI	8.606	0.228	0.950	42.173	15.355	47.908
		9.248	0.263	0.958	47.518	17.886	53.025
First rib	Median	9.409	0.275	0.953	48.631	24.411	52.194
(m = 2)	95% CI	9.067	0.255	0.950	46.011	22.362	50.048
		9.751	0.296	0.958	50.585	26.334	53.906
Pubic symphysis	Median	10.898	0.370	0.953	53.814	27.707	62.436
(m = 3)	95% CI	10.436	0.343	0.948	51.151	25.107	59.714
		11.315	0.398	0.963	55.553	29.902	64.387
Sacroiliac complex	Median	8.438	0.220	0.950	44.238	20.694	49.290
(m = 4)	95% CI	8.075	0.200	0.950	41.626	18.741	47.493
		8.741	0.239	0.953	47.091	22.445	52.320
Acetabulum	Median	8.833	0.229	0.950	40.223	31.121	44.259
(m = 3)	95% CI	8.490	0.210	0.950	38.257	29.463	41.867
		9.116	0.247	0.953	42.034	32.687	46.304
Degenerative traits	Median	6.929	0.147	0.950	32.126	29.464	33.763
(m = 39)	95% CI	6.694	0.133	0.950	30.667	28.096	31.766
		7.154	0.157	0.953	33.471	30.717	35.619
Standard	Median	6.561	0.145	0.950	34.142	13.397	43.611
(m = 16)	95% CI	6.277	0.132	0.950	32.354	12.340	41.685
		6.855	0.157	0.950	35.794	14.618	45.507
Full	Median	5.899	0.118	0.950	29.471	16.316	37.617
(m = 64)	95% CI	5.677	0.110	0.950	27.857	15.076	35.629
		6.121	0.127	0.950	31.004	17.534	39.204

Table C.3 Leave-one-out cross-validation for models built on different fractions of available skeletal traits. Ensembled deep randomized neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.942	0.121	0.953	30.276	15.745	36.278
(m≈57)	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%	Median	5.970	0.122	0.953	30.476	15.941	36.332
(m≈51)	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%	Median	6.028	0.124	0.953	30.711	16.182	36.518
(m ≈ 44)	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%	Median	6.078	0.125	0.953	30.975	16.342	36.716
(m ≈ 38)	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%	Median	6.173	0.128	0.953	31.502	16.684	37.040
(m ≈ 32)	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%	Median	6.305	0.132	0.953	32.146	17.153	37.511
(m ≈ 25)	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%	Median	6.501	0.138	0.953	33.097	17.923	38.203
(m ≈ 19)	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%	Median	6.957	0.154	0.953	35.321	19.986	39.742
(m ≈ 12)	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%	Median	7.952	0.192	0.955	39.733	26.846	43.076
(m ≈ 6)	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

Table C.4 Leave-one-out cross-validation for models built on different fractions of available skeletal traits. Ensembled deep randomized neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90% (m≈57)	Median	5.942	0.121	0.950	29.601	16.510	37.581
	95% CI	5.699	0.110	0.950	27.753	15.032	35.563
		6.198	0.131	0.950	31.570	18.860	39.823
80% (m≈51)	Median	5.970	0.122	0.950	29.798	16.694	37.668
	95% CI	5.702	0.108	0.950	27.990	14.920	35.535
		6.235	0.132	0.950	31.590	19.326	39.776
70% (m ≈ 44)	Median	6.028	0.124	0.950	30.003	16.914	37.823
	95% CI	5.737	0.108	0.950	28.075	14.738	35.561
		6.376	0.137	0.950	32.114	20.557	40.405
60% (m ≈ 38)	Median	6.078	0.125	0.950	30.278	17.128	38.025
	95% CI	5.768	0.108	0.950	28.188	14.560	35.652
		6.441	0.140	0.950	32.774	21.483	40.739
50% (m ≈ 32)	Median	6.173	0.128	0.950	30.812	17.440	38.352
	95% CI	5.819	0.111	0.950	28.521	14.304	35.647
		6.648	0.146	0.950	33.287	23.025	41.491
40% (m ≈ 25)	Median	6.305	0.132	0.950	31.448	17.841	38.889
	95% CI	5.903	0.114	0.950	28.960	14.420	35.674
		6.797	0.153	0.950	34.142	24.051	42.231
30% (m ≈ 19)	Median	6.501	0.138	0.950	32.418	18.662	39.584
	95% CI	6.046	0.118	0.950	29.520	14.345	36.119
		7.096	0.163	0.950	35.162	26.031	43.309
20% (m ≈ 12)	Median	6.957	0.154	0.950	34.370	20.500	41.038
	95% CI	6.316	0.127	0.950	30.927	14.348	36.011
		7.674	0.184	0.950	37.852	29.314	45.647
10% (m ≈ 6)	Median	7.952	0.192	0.950	38.472	27.257	43.532
	95% CI	6.968	0.154	0.950	34.134	15.510	37.887
		9.214	0.256	0.953	45.233	34.819	50.951

Table C.5 Leave-one-out cross-validation for models built on pre-specified skeletal traits sets. Deep supervised autoencoder neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.212	0.420	0.935	63.254	41.679	69.192
(m = 9)	95% CI	14.441	0.386	0.922	61.130	38.004	67.221
		15.849	0.452	0.950	65.373	45.298	70.715
Axial	Median	7.952	0.107	0.948	36.885	31.182	39.611
(m = 16)	95% CI	7.672	0.098	0.938	35.390	29.738	37.998
		8.219	0.115	0.958	38.212	32.462	41.199
Appendicular	Median	7.482	0.095	0.955	36.240	27.542	38.927
(m = 16)	95% CI	7.168	0.087	0.943	34.711	26.168	36.962
		7.771	0.102	0.968	37.757	28.804	40.691
Clavicle	Median	9.087	0.144	0.955	48.944	16.148	49.652
(m = 2)	95% CI	8.722	0.132	0.948	47.065	14.905	48.265
		9.439	0.155	0.965	50.484	17.400	51.137
First rib	Median	9.577	0.163	0.948	47.030	21.697	49.514
(m = 2)	95% CI	9.197	0.150	0.938	45.454	20.240	47.667
		9.925	0.175	0.960	48.516	23.141	51.258
Pubic symphysis	Median	11.407	0.225	0.940	53.778	27.069	56.562
(m = 3)	95% CI	10.870	0.207	0.927	51.190	24.896	54.734
		11.873	0.243	0.958	55.933	29.216	58.294
Sacroiliac complex	Median	8.288	0.125	0.932	43.828	17.542	46.645
(m = 4)	95% CI	7.936	0.115	0.920	41.767	16.001	44.883
		8.642	0.136	0.948	45.615	19.008	48.377
Acetabulum	Median	8.705	0.129	0.950	41.210	28.603	45.112
(m = 3)	95% CI	8.396	0.119	0.943	39.642	26.936	43.423
		9.005	0.139	0.958	42.654	30.032	46.666
Degenerative traits	Median	6.695	0.075	0.949	32.304	26.748	33.958
(m = 39)	95% CI	6.430	0.069	0.935	31.101	25.380	32.186
		6.957	0.081	0.963	33.489	28.017	35.681
Standard	Median	6.415	0.080	0.940	33.123	11.246	39.482
(m = 16)	95% CI	6.104	0.072	0.925	31.121	10.221	37.638
		6.703	0.087	0.955	35.020	12.228	41.197
Full	Median	5.608	0.059	0.935	28.379	12.494	34.557
(m = 64)	95% CI	5.352	0.054	0.920	26.822	11.105	32.803
		5.860	0.065	0.950	29.938	14.556	36.330

Table C.6 Leave-one-out cross-validation for models built on pre-specified skeletal traits sets. Deep supervised autoencoder neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
Sutures	Median	15.212	0.420	0.950	68.174	40.219	80.033
(m = 9)	95% CI	14.441	0.386	0.948	65.044	36.130	75.371
		15.849	0.452	0.953	71.329	44.194	84.649
Axial	Median	7.952	0.107	0.950	38.611	32.188	40.308
(m = 16)	95% CI	7.672	0.098	0.950	36.411	30.593	37.723
		8.219	0.115	0.953	40.458	33.629	43.011
Appendicular	Median	7.482	0.095	0.950	36.239	28.088	39.281
(m = 16)	95% CI	7.168	0.087	0.950	33.997	26.610	36.526
		7.771	0.102	0.953	38.388	29.478	41.895
Clavicle	Median	9.087	0.144	0.950	46.713	14.342	51.311
(m = 2)	95% CI	8.722	0.132	0.950	43.736	13.100	48.781
		9.439	0.155	0.955	49.990	15.896	53.248
First rib	Median	9.577	0.163	0.950	49.305	20.798	50.406
(m = 2)	95% CI	9.197	0.150	0.950	47.144	19.089	47.811
		9.925	0.175	0.955	50.252	22.462	52.299
Pubic symphysis	Median	11.407	0.225	0.953	55.966	24.752	60.064
(m = 3)	95% CI	10.870	0.207	0.950	53.473	22.776	57.695
		11.873	0.243	0.965	58.335	26.671	62.889
Sacroiliac complex	Median	8.288	0.125	0.950	46.881	17.920	51.686
(m = 4)	95% CI	7.936	0.115	0.950	44.394	16.231	49.305
		8.642	0.136	0.955	48.729	19.625	54.145
Acetabulum	Median	8.705	0.129	0.950	42.203	29.043	46.478
(m = 3)	95% CI	8.396	0.119	0.950	40.607	27.062	44.050
		9.005	0.139	0.953	44.410	30.487	49.676
Degenerative traits	Median	6.695	0.075	0.950	32.836	27.895	34.575
(m = 39)	95% CI	6.430	0.069	0.950	31.086	26.335	32.327
		6.957	0.081	0.953	34.485	29.176	36.753
Standard	Median	6.415	0.080	0.950	33.844	11.329	42.738
(m = 16)	95% CI	6.104	0.072	0.950	31.973	10.209	40.596
		6.703	0.087	0.950	36.055	12.608	45.014
Full	Median	5.608	0.059	0.950	29.860	13.160	37.951
(m = 64)	95% CI	5.352	0.054	0.950	27.901	11.482	35.837
		5.860	0.065	0.950	31.894	15.439	40.252

Table C.7 Leave-one-out cross-validation for models built on different fractions of available skeletal traits. Deep supervised autoencoder neural network with truncated gaussian regression uncertainty model.

				Truncated Gaussian			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.653	0.060	0.938	28.657	12.746	34.707
(m≈57)	95% CI	5.398	0.055	0.920	27.120	11.210	32.985
		5.931	0.066	0.953	30.290	15.325	36.428
80%	Median	5.706	0.061	0.938	28.887	12.884	34.909
(m≈51)	95% CI	5.428	0.055	0.920	27.265	11.167	32.940
		6.044	0.068	0.953	31.056	15.922	36.906
70%	Median	5.770	0.063	0.938	29.251	13.368	35.057
(m ≈ 44)	95% CI	5.443	0.055	0.917	27.310	11.219	32.963
		6.109	0.069	0.953	31.247	16.897	37.186
60%	Median	5.827	0.064	0.938	29.626	13.647	35.271
(m ≈ 38)	95% CI	5.527	0.057	0.920	27.819	11.298	33.244
		6.217	0.071	0.953	31.645	18.256	37.271
50%	Median	5.917	0.065	0.938	29.976	14.021	35.619
(m ≈ 32)	95% CI	5.591	0.058	0.920	28.113	11.418	33.463
		6.392	0.074	0.955	32.397	19.069	37.987
40%	Median	6.091	0.069	0.939	30.984	14.794	36.182
(m ≈ 25)	95% CI	5.694	0.060	0.922	28.694	11.718	33.743
		6.592	0.078	0.955	33.412	21.548	38.515
30%	Median	6.269	0.073	0.940	31.834	15.516	36.829
(m ≈ 19)	95% CI	5.849	0.063	0.922	29.579	11.937	34.292
		6.918	0.085	0.958	34.678	23.101	39.827
20%	Median	6.740	0.082	0.943	33.982	17.049	38.323
(m ≈ 12)	95% CI	6.100	0.069	0.925	30.890	12.451	35.218
		7.510	0.099	0.960	37.552	27.047	42.111
10%	Median	7.777	0.105	0.945	38.169	23.297	41.827
(m ≈ 6)	95% CI	6.771	0.083	0.927	34.197	13.519	37.846
		9.044	0.142	0.963	44.295	31.549	47.740

Table C.8 Leave-one-out cross-validation for models built on different fractions of available skeletal traits. Deep supervised autoencoder neural network with conformal prediction regression uncertainty model.

				Conformal Prediction			
		Accuracy	Bias	Validity	Efficiency		
Available Traits (%)		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95%CI	
90%	Median	5.653	0.060	0.950	30.108	13.431	38.132
(m≈57)	95% CI	5.398	0.055	0.950	28.011	11.556	35.709
		5.931	0.066	0.950	32.303	16.140	40.459
80%	Median	5.706	0.061	0.950	30.336	13.534	38.252
(m≈51)	95% CI	5.428	0.055	0.950	28.165	11.499	35.852
		6.044	0.068	0.950	32.641	16.905	40.973
70%	Median	5.770	0.063	0.950	30.678	14.042	38.378
(m ≈ 44)	95% CI	5.443	0.055	0.950	28.388	11.509	35.968
		6.109	0.069	0.950	33.115	18.086	41.075
60%	Median	5.827	0.064	0.950	30.987	14.377	38.634
(m ≈ 38)	95% CI	5.527	0.057	0.950	28.507	11.576	35.804
		6.217	0.071	0.950	33.477	19.430	41.418
50%	Median	5.917	0.065	0.950	31.330	14.708	38.856
(m ≈ 32)	95% CI	5.591	0.058	0.950	28.922	11.702	36.137
		6.392	0.074	0.950	34.035	20.015	41.977
40%	Median	6.091	0.069	0.950	32.203	15.454	39.393
(m ≈ 25)	95% CI	5.694	0.060	0.950	29.456	12.019	36.321
		6.592	0.078	0.950	35.090	22.724	42.666
30%	Median	6.269	0.073	0.950	32.996	16.240	39.973
(m ≈ 19)	95% CI	5.849	0.063	0.950	30.219	12.029	36.413
		6.918	0.085	0.950	36.480	24.039	43.923
20%	Median	6.740	0.082	0.950	34.984	17.642	41.255
(m ≈ 12)	95% CI	6.100	0.069	0.950	31.600	12.503	36.785
		7.510	0.099	0.953	38.860	27.961	45.996
10%	Median	7.777	0.105	0.950	39.329	24.062	44.424
(m ≈ 6)	95% CI	6.771	0.083	0.950	34.259	13.154	38.648
		9.044	0.142	0.953	45.752	32.337	51.567