



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Modeling Surface Morphology of the Pubic Symphysis

Author(s): Dennis E Slice, Jieun Kim, Detelina K. Stoyanova, Cristina Figueroa-Soto, Bridget F. B. Algee-Hewitt

Document Number: 255134

Date Received: August 2020

Award Number: 2015-DN-BX-K010

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

FINAL REPORT

MODELING SURFACE MORPHOLOGY OF THE PUBIC SYMPHYSIS

AWARD NO: 2015-DN-BX-K010

AUTHORS

Dennis E Slice (PI), Jieun Kim (Postdoc), Detelina K. Stoyanova (Postdoc),
Cristina Figueroa-Soto (Doctoral student), Bridget F. B. Algee-Hewitt (Co-PI)

Dennis E. Slice, PhD.

Department of Scientific Computing
400 Dirac Science Library
Florida State University
Tallahassee, FL, USA 32306-4120.
Email: dslice@fsu.edu
Phone: 850-644-1010

Jieun Kim, PhD.

Department of Scientific Computing
400 Dirac Science Library
Florida State University
Tallahassee, FL, USA 32306-4120.
Email: jkim17@fsu.edu
Phone: 850-644-1010

Detelina K. Stoyanova, PhD.

Department of Scientific Computing
400 Dirac Science Library
Florida State University
Tallahassee, FL, USA 32306-4120.
Email: dks10d@fsu.edu
Phone: 850-644-1010

Cristina Figueroa-Soto, MA.

The University of Tennessee
Department of Anthropology
Knoxville, TN 37996.
Email: cfiguer1@vols.utk.edu
Phone: 865-974-4408

Bridget F. B. Algee-Hewitt, PhD.

Center for Comparative Studies
in Race & Ethnicity
Stanford University
450 Serra Mall, Building 360
Stanford, CA, 94305.
Email: bridgeta@stanford.edu
Tel: 865-382-7244

TABLE OF CONTENTS

1.ABSTRACT.....	4
2.EXECUTIVE SUMMARY.....	4
2.1.Synopsis of the Problem.....	4
2.2.Purpose of this Project.....	7
2.3.Research Design.....	8
2.4.Findings, Conclusions, and Impact.....	9
3.INTRODUCTION.....	12
3.1.Statement of the Problem.....	12
3.2.Literature Citations and Review.....	12
3.3.Statement of Hypothesis or Rationale for the Research.....	13
4.METHODS.....	14
4.1.Data Collection.....	14
4.2.Age-At-Death Estimation Methods.....	16
5.RESULTS.....	17
5.1.Statement of the Results.....	17
6.CONCLUSION.....	21
6.1.Discussion of Findings.....	21
6.2.Discussion of Implications for Policy and Practice.....	24
6.3.Discussion of Future Work.....	26
6.4.Tables.....	28
6.5.Figures.....	29
7.REFERENCES.....	39
8.DISSEMINATION OF RESEARCH FINDINGS.....	45

1. ABSTRACT

Age-at-death is one among the key parameters of interest when building a biological profile from an unknown adult skeleton. The pubic symphysis is said to be the preferred and most commonly used skeletal element for age-at-death estimation. The conventional approach to age estimation from this indicator requires the visual assessment of the bone morphology and its association with predefined age-ranges. While this practice is appealing for its simplicity, the process of bone-to-phase matching is methodologically subjective: it is conditional upon the range of variation expressed in the comparative sample, or the typicality of the skeletal element under analysis, and the macromorphoscopic assessment is driven by the experience and training of the practitioner. Problems of method and observer error are well established, yet efforts to make improvements to these approaches have been few and have gained little traction since their introduction. We have argued, therefore, that there is an immediate need to develop alternative tools for the rigorous quantification of age-related skeletal changes in the pubic symphysis. Further, we have advocated for greater diversity in the populations for which methods have been developed and on which they have been validated, especially with respect to the present demographic composition of the United States. We have developed a fully computational method that employs three shape measures of the pubic symphysis that are associated with the age-at-death. We have built and tested models for a wide sampling of modern American and Asian populations. We demonstrate that this approach approximates true chronological age with moderate to high accuracy, regardless of side sampling (asymmetry) and level of training (observer error). We have also generated results suggestive of the need for population-specific methods. To implement this computational approach, we have developed the free, open-source software, *forAge*, including sample data scans, to ensure the easy and consistent implementation of our methods by a wide variety of forensic caseworkers. In reducing methodological subjectivity, providing fully quantified estimates and rates of error, and facilitating the consistent implantation of this method via our project-specific software, we contend that this approach better meets medico-legal standards of scientific evidence.

2. EXECUTIVE SUMMARY

2.1. Synopsis of the Problem.

The estimation of chronological age-at-death from the adult skeleton is a key component of the biological profile produced by forensic anthropologists.

Knowing age is essential for reducing the investigative burden by reducing the list of potential victims in the identification process: too wide an interval, irrelevant cases are included, while too limited a range, potential matches are lost. Further, age-at-death is often a prerequisite for the accurate interpretation of other skeletal signatures, like pathological conditions (Milner and Boldsen, 2017; Milner et al, 2008; Ortner, 2003), that may give valuable clues to an individual's identity, increasing the likelihood that the remains represent a specific person (Konigsberg et al, 2008; Steadman et al, 2006). In the medico-legal context, it is crucial, therefore, that estimates produced at the level of the forensic case are both precise and accurate, such that the difference between inferred and true age is small and width of the confidence interval is narrow. Because there exists an imperfect correlation between chronological age, simply defined by calendar years since birth, and physiological age, inferred from the state of morphological expression of degenerative indicators that change sequentially with time, some quantity of error in estimation is unavoidable (Kemkes-Grottenthaler, 2002; Konigsberg, 2015; Nawrocki, 2010). This error can be attributed to the life-history-driven variation in each skeleton and is said to result its own unquantifiable degree of error (Boldsen et al, 2002). It can also result from methodological limitations, whether introduced at the stages of technique development or implementation, both of which include some combination of data collection, indicator evaluation, and age determination.

In this project, we have focused on this latter, technique-driven, error as it can be experimentally controlled, tested, and quantified. In common practice, forensic anthropologists estimate adult age-at-death by quantifying the morphological variation in a skeletal indicator associated with aging in terms of a suite of grossly observable characteristics, using an ordinal categorical scoring system. The pubic symphysis is the most widely exploited, and preferred, of the skeletal indicators for adult age estimation (Garvin and Passalacqua, 2012). In this case, the morphology of the symphyseal surface in question is compared against a set of sex-specific criteria that represent a series of pre-defined phases. The case-specific age-at-death is then estimated from the age-range previously associated with the assigned phase (Brooks and Suchey 1990). Other well-known skeletal age indicators are distributed throughout the skeleton, and the most familiar include the auricular surfaces, ribs, and clavicles. They also undergo the same kind of sequential, degenerative alteration (in shape, size, and texture) to bone morphology with increasing age. They are, accordingly, subject to similar visual scoring criteria

and equally permit the inference of chronological age from the state of their observed biological features (Buckberry and Chamberlain, 2002; DiGangi et al, 2009; Falys et al, 2006; Hartnett, 2010; Igarashi et al, 2005; Işcan, 1987; Işcan et al, 1984; Işcan et al, 1985; Langley-Shirley and Jantz, 2010; Lovejoy et al, 1985; McKern and Stewart, 1957; Merritt, 2014; Murray and Murray, 1991; Osborne et al, 2004; San Millán et al, 2013). These “macromorphoscopic” approaches are understandably popular in forensic casework as they are well-established in the literature of biological anthropology, giving the practitioner the scientific precedent needed for medical-legal reporting. In practice, they are also very appealing as they are simple to apply and place little demands on the practitioner with regards to equipment, space for evaluation, and statistical analysis. Yet, such bone-to-phase matching techniques are at the same time problematic for the fact that they are conditional upon the range of variation captured in the comparative reference samples and their success in application is largely driven by the user. More specifically, their efficacy is dependent upon both the morphological typicality of the skeletal elements under analysis relative to the method-specific standards (e.g., text-based descriptions, photographs, and reference casts) as well as the training and experience (i.e., in source and degree) of any given practitioner. Accordingly, the standard indicator assessment methods are often quite subjective: a point already well made by the body of literature on the problems of method and observer error in the conventional approaches to skeletal age estimation (Fojas et al. 2015; Hoppa 2000; Kim J et al. 2018; Kimmerle et al. 2008a; Kimmerle et al. 2008b; Saunders et al. 1992; Shirley and Ramirez Montes 2015).

The comparative nature of these methods also imposes other limitations on the bone feature evaluation and the final prediction of age. Except for those techniques that adopt an explicitly component-based approach to feature scoring, there are no options for deriving ages from indicators with areas that are fragmented, damaged, or obscured. Further, the foundational methods focus on the analysis of a single indicator. They ignore the valuable information on age that may be sourced from the contribution of many, large- and small-scale changes to the skeleton. This single indicator approach places on the individual practitioner the burden of determining if and how multiple indicators, distributed throughout the available skeletal elements and scored independently, should be combined to produce a single, more inclusive age estimate for a given case and how to contend with instances when different indicators yield conflicting results. This system necessarily encourages either

the loss of information by disregarding all other/some indicator data or the proliferation of *ad hoc* approaches to a “gestalt” age analysis, thereby producing estimates of age that are again practitioner dependent.

Finally, many of the prevailing methods have been developed and validated on only a subset of the world populations, despite the fact that uncharacterized skeletal cases can vary in time, space, and life history. Preference is typically given those groups for which skeletal reference materials is readily available. Forensic anthropologists must contend with the fact that unidentified skeletal remains rarely derive from closed populations, as in select mass fatality scenarios such as the American KIA repatriation, the New York City World Trade Center attacks (Adams and Byrd 2008), the recent sunken ferry, Sewol, incident in South Korea and the Branch Davidian compound recovery (Ubelaker et al, 1995), and most often represent unknown source populations. Forensic anthropologists also routinely engage casework outside the geographic boundaries of North America (Djurić et al, 2007a; Djurić et al, 2007b; Hens and Belcastro, 2012; Hens et al, 2008; Rissech et al, 2012; Schmitt, 2004). Even at home, practitioners in the United States must account for continued diversification of the American population-at-large in its composition, the presence of secular change, and shifts in the demographics of those persons who make up the present pool of local forensic cases (Shrestha, 2011; Frey, 2015; Taylor, 2014; Algee-Hewitt et al, 2018; Algee-Hewitt, 2017a; Algee-Hewitt, 2017b; Hughes et al, 2017; Jantz and Jantz, 1999; Klales, 2016; Klepinger, 2001; Komar and Grivas, 2008; Spradley, et al. 2008; Wescott and Jantz, 2005).

2.2. Purpose of this Project.

This project, supported by the National Institute of Justice (NIJ award 2015-DN-BX-K013), was motivated by our belief that, despite these known limitations in methodology, improvement in skeletal age estimation can be achieved. To this point, we have argued that transformative work in age-at-death estimation should seek, first, to identify the potential sources of error in the current gross-assessment methodologies and, next, to offer computational solutions to the prediction of chronological age. This project was developed, therefore, with the purpose to (1) draw upon technological innovations in semi-to-fully-automated data capture technologies and methodologies of analysis to identify, ideally new and/or improved, measures of skeletal morphology that are highly correlated with chronological age, (2) consider the effects of bilateral asymmetry and define its relationship to individual life-history factors, like

body size (weight/BMI/height) and parturition, when choosing between the left or right age or both indicator(s), (3) account for confounding factors like fragmentation and sex differences at the stages of data acquisition and statistical inference of age, (4) produce methods that are either population-specific, and uniquely suitable for the diverse peoples that are represented in contemporary casework, or invariant to such biogeographic variation, and (5) generate age-at-death estimates with high precision, accuracy, and repeatability that offer reliable forensic evidence that meet heightened standards of medico-legal admissibility, including reported rates of method and observer error at all stages of analysis – from data collection to age computation. In order to ensure that the computational framework developed over the course of this project would be accessible to the forensic community, when broadly defined and regardless of the practitioner's degree of osteological, 3D laser scanning, image processing, or statistical training, we sought to produce practical guidelines for the correct implementation of our proposed methodologies and a free, open-source software package for the user-friendly execution of our shape-based algorithms and regression equations as well as the reliable and repeatable (re)production of age estimates using our semi-automated framework and built-in database of reference scans.

2.3. Research Design

In keeping with the mission of the National Institute of Justice award, to serve immediate forensic science needs, our research design was structured to deliver practical methodologies, user guidelines, computing tools and robust results that are rooted in sound biological and statistical theory and meet the complex demands of forensic anthropological casework in the current medico-legal setting. However, in the spirit of sustainability, broader utility, and cross-disciplinary impact, we have worked to design a project whose deliverables are of value to members of the forensic and biological anthropological research, teaching, and practitioner communities alike. Accordingly, we have developed a data acquisition protocol (for scan settings/seg-up, editing, data exporting) that is generally applicable to other skeletal elements and a set of computational tools (including our published age estimation equations, shape algorithms, and comprehensive *forAge* software) that can be extended to other skeletal indicators, whether to assess age or to quantify the bone features shapes for their subsequent association with, for examples, other parameters of personal identity, like sex and ancestry, or different taphonomic, like pathological, traumatic, and destructive, conditions. We have also adopted a multi-population approach to our method development, such that our framework

has the potential to be implemented in various forensic analysis and biological research settings, at home or abroad, where the demographic composition of the cases or samples is diverse. To this point, we have pursued our data collection, tested our preliminary single-sex/population models, and, as needed, re-built our equations using samples representing males and females from many different groups from the Americas, Europe and Asia, representing such socio-geographic identifiers as Black, White, Hispanic (Spanish, Puerto-Rican, and Mexican), Japanese, Korean and Thai. Further, we have extended our techniques for use on archaeological collections, validating our work with scans collected on the PaleoIndian sample from Windover Pond (<https://fsu.digital.flvc.org/islandora/object/fsu%3Awindover>).

2.4. Findings, Conclusions, and Impact.

Our project has sought to respond to the present methodological limitations in age-at-death estimation by addressing areas of potential improvement: we focus, specifically on reducing subjectivity and error, when considering the most commonly used element in “aging the skeleton” as our test indicator. To this goal, we have designed a fully computational framework for adult age-at-death estimation from the pubic symphysis. We drew upon our preliminary laser-scan work, first published in Slice and Algee-Hewitt (2015), which advanced a novel variance-based approach to inferring chronological age from a surface sampling of the pubic symphysis face using principal component and regression methods. First, building upon these findings, we have offered methodological improvement in data acquisition by delivering (1) new three-dimensional laser-scan data on, primarily, the pubic symphysis, but also including test samples of other classic age indicators (medial surface of the clavicle, sternal surface of the rib, and auricular surface of the os coxa), (2) creating the first published protocols for (a) collecting x,y,z coordinate data from said skeletal elements using a portable, low-cost 3D laser scan system, (b) editing the resulting scans for regions of interest, and (c) processing the numeric data for downstream visual and statistical manipulation. Second, we have improved data analysis and age inference by (1) refining the *SAH-score* previously reported in Slice and Algee-Hewitt (2015), (2) implementing shape-analysis algorithms to produce two new measures, the *Bending Energy (BE)* and *Ventral Curvature (VC) scores*, to capture the variation in the bone surface as it undergoes morphological evolution — whether through the build-up or break-down of bone — over the course of physiological aging, and (3) developing new statistical, regression-based, models for the prediction of chronological age-at-death that are (a) built to use each of the three shape

scores or two of these scores combined, and (b) applicable to diverse populations, including both sexes and various bio-social, geographic, and ancestry groups. Second, we have built a substantial database of over 1200 pubic symphysis scans obtained on documented contemporary individuals of self-reported White, Black, Hispanic/Latinx in the United States, known individuals sampled from modern Mexican, Puerto Rican, Japanese, Thai, and Spanish populations, and three collections of cases representing persons of unknown identity, one representing undocumented individuals of current forensic significance recovered along the US-México Border and another representing young, male Korean soldiers, and archaeological remains from the PaleoIndian site of Windover Pond. Finally, we have made publicly available for free download a multi-platform, open-source, software package, *forAge*. Written specifically for this project, *forAge* allows the user to easily (1) upload their own data as a simple text file of coordinate values, (2) implement our shape algorithms to generate the SAH, BE and VC scores, and (3) use one or a combination of these shape scores in regression-based analyses based on our reference database of pubic symphyseal scans to produce ages-at-death for their target samples.

We have convincingly shown how, by exploiting coordinate data that is quickly and easily derived from low-cost and portable desktop scanners, it is possible to design a semi-automated framework under which we can estimate age-at-death with good accuracy and without the subjectivity that accompanies a traditional morphoscopic assessment (Slice and Algee-Hewitt, 2015, Stoyanova et al, 2015, Stoyanova et al, 2017, Kim et al, 2018, Stoyanova et al, 2018). The shape-based metrics generated by our published techniques show significant linear associations with chronological age, yielding moderate to highly positive correlation coefficients. Our regression equations provide consistent estimates that fall within useful age intervals and allow for the quantification of method error, by computing, for example, RMSE, R^2 and p-values statistics for model evaluation, agreement metrics for assessing observer error, and individual age ranges for measuring uncertainty in case-specific analyses. Our method output – and more, specifically, the three shape scores – can also be used to reveal trends in morphological variation associated with life-history factors, asymmetry in bilateral indicators, and different populations of origin. As key project deliverables, our published papers using these scores can serve as the foundation for recommendations on best practices related to not only computational age-at-death estimation using three-dimensional shape data but also the choice of sides (left or right or both) for feature assessment, the

need to include covariates, such as BMI or stature, in the age estimation models, and the importance of accounting for spatial, temporal, sex and population-specific differences in the skeletal expression of the senescent process when choosing among available reference samples and the statistical equations used in the prediction of chronological age. Finally, our extension work on other skeletal indicators has offered promising results, indicating, in particular, that our current data collection protocols and computations for the three published shape scores can be translated to the auricular surface with minimal refinement required.

We anticipate exploiting not only this potential for advancing new work that follows along the trajectory of the research completed under this NIJ project, but also expanding the scope of our computational approach to alternative techniques that capitalize upon the advance in artificial intelligence and machine learning (AI-ML) algorithms, which have been successfully used for shape analysis and shape pattern recognition (Chen et al, 2016; Liang et al, 2017; Lim et al, 2016; Omodaka et al, 2018; Wade et al, 2017).

This project has important implications for policy and practice in Forensic Anthropology as it has demonstrated the potential to transform the way that skeletal assessment and age determination is performed by advancing a “fully-computational framework” for age-at-death estimation. Specifically, it brings to the field a more objective method for (a) documenting surface morphology of classic skeletal age indicators, (b) quantifying age-related changes in these skeletal indicators, (c) reducing method and observer-related error, and (d) providing robust, repeatable statistical results for chronological age-at-death in cases distributed across the adult life-span and sourced from various populations. When these factors are considered together, this project provides a comprehensive methodology that better meets the medico-legal demands for standards of forensic evidence, as they have been made clear in the guidelines by the Scientific Working Group for Forensic Anthropology (SWGANTH 2013) and the National Research Council (National Research Council 2009) and which are applicable to forensic anthropological case analyses. Owing to its semi-automated and fully-computational design, this project’s methodological framework can be implemented by practitioners with varying kinds of expertise and levels of experience. With the use of the accompanying software, *forAge*, our shape and age-prediction analyses can be easily performed by any researcher in the absence of the actual skeletal material, with minimal computational resources, and without the need to personally write code for,

and execute, the statistical procedures. The flexible design of our methodology allows for the straightforward incorporation of new data from scans representing other populations (e.g., self-reported group identifiers, geographic ancestries, sexes), different temporal periods (i.e., to address secular change or assess forensic v. archaeological samples), paired elements (i.e., to account for asymmetry), and even different age-informative indicators (e.g., auricular surfaces, ribs, clavicles; dental wear). The forward-thinking approach to quantifying changes in shape computationally that underlies this project's age estimation methods will permit future development, refinement, and expansion. This project's work also has the potential to impact other areas of research, inspiring new methods for and initiating improvements to parameter estimation for the biological profile in Forensic Anthropology.

3. INTRODUCTION

3.1. Statement of the Problem

Age-at-death estimation is an important component in the construction of a biological profile by forensic anthropologists. Yet, standard, widely-used methods for this tend to be subjective with varying and inconsistent results based on user experience and fail to meet requisite legal standards for scientific evidence. This project sought to address these issues by developing and testing new, objective, and computational methods for age-at-death estimation using laser scans of the pubic symphysis, constructing an extensive database of such scans to better represent U. S. and global populations, and providing free software and training in the use of these methods.

3.2. Literature Citations and Review

The pubic symphysis (Figure 1) is the most widely used, and the most preferred, skeletal indicator when estimating age-at-death for adult human remains (Garvin and Passalacqua 2012). Standard practice requires the visual comparison of the bone surface morphology of the pubic symphysis against a set of population-specific criteria that represent a series of pre-defined phases. The case-specific age-at-death is estimated from the age-range previously associated with the assigned phase (Brooks and Suchey 1990). While this approach is appealing for its simplicity, the process of bone-to-phase matching is conditional upon the variation in the comparative samples and driven by the user; as such, it is methodologically subjective. The efficacy of this kind of approach is, therefore, dependent upon both the nature of training and practical experience of the practitioner as well as the morphological typicality of the skeletal elements under analysis relative to the method-specific standards,

e.g., text-based descriptions, photographs, and comparative casts. The well-known problems of method and observer error in skeletal age estimation have suggested an immediate and ongoing need to develop alternative tools for the rigorous quantification of age-related change in skeletal morphology that better approximate true chronological age, regardless of sample source (population of origin or sex) and level of practitioner training or experience, and better meet medico-legal standards of scientific evidence (Ritz-Timme et al, 2000; Algee-Hewitt, 2011). Moreover, as forensic anthropologists routinely engage casework outside the geographic boundaries of the United States and as the American population continues to diversify in its composition, it is critical that efforts are made to reach a clear understanding of age-related changes in under-studied populations and, in turn, to develop standardized methods that can be equally applied across biogeographic groups on a global scale.

3.3. Statement of Hypothesis or Rationale for the Research

This project (2015-DN-BX-K010) sought to address the limitations of more subjective methods of age-at-death estimation by:

1. Testing, extending, and developing objective, computational methods for age-at-death estimation.
2. Constructing a database of surface scans of the pubic symphysis which could be used to test and develop objective, computational methods for age-at-death determination and provide better representation of populations in the U. S. and around the world.
3. Delivering free, open-source software for the implementation of objective, computational methods for age-at-death determination.

The work of Slice and Algee-Hewitt (2015) demonstrated an objective method for age-at-death estimations based on the overall variability of the pubic symphysis surface. For this project, we sought to build upon this to develop a new method to incorporate not only the magnitude of variation in the surface of the pubic symphysis, but also its spatial organization (the thin-plate spline or bending-energy method). Furthermore, experience in working with these methods and the source material was to be used to devise and test other methods resulting in work on the ventral-curvature method of age-at-death estimation.

Methods can, and arguably should, be developed on smaller, more homogeneous samples to identify effect size and eliminate controllable sources of extraneous variation, such as sex or ancestry. However, they must ultimately be examined for potential effects of such factors. In the course of

this research, we examined the performance of the developed methods with respect to a number of factors that could impinge upon their efficacy. These included sex, asymmetry, ancestry, user-reliability, and life-history factors such as stature, weight, or body-mass index.

To provide a sufficient source of data for the above research, laser scans of pubic symphyses were collected from forensic and medical collections representative of U. S. population subgroups. Similar scans were also obtained from collections in Japan, Thailand, and Korea. While not originally anticipated, it was possible to support the further extension of this database to include more extensive Hispanic data from the Mexico and Puerto Rico, and data were included from a Paleoindian collection to better understand the benefits and limits of the new methods.

We further took advantage of the data-collection trips oriented toward the scanning of pubic symphyses to collect other data for which our new methods might be applicable or for which new methods might be developed. This includes the auricular surface, the sternal end of the clavicle, and sternal 4th rib end.

The results of the above were disseminated through poster, presentations, and peer-reviewed publications, and training in the methods and software were provided to the community of potential end-users through workshops, symposia, and online resources.

4. METHODS

4.1. Data Collection

Data Collection Protocol: Laser scans of contemporary material were made using the NEXTENGINE 3D Desktop Scanner, such that the selected os coxa was captured over three consecutive angles. The pubic symphysis was positioned within the center scan, with an offset scan to both its ventral and dorsal sides (Figure 1). The three resulting scans were grouped as a family, aligned, and fused such that a multidimensional model (i.e., a coordinate mesh with a photometric surface texture map) of the full bone was generated. The face of the pubic symphysis was then isolated within this full model and the surrounding areas were deleted, leaving only the trimmed surface of the pubic symphysis. The trimmed mesh was converted to either a Wavefront geometric object file (OBJ) or Polygon file (PLY) and stored in ASCII format. The data analyzed here are, therefore, the three-dimensional coordinates of points (x, y, z values) that characterize the surface of the pubic symphysis for each scanned individual and that have been saved in a .obj (or .ply) file. The procedure, as

just described, was also followed to obtain coordinate data for the Suchey-Brooks and McKern and Steward casts.

All scans were taken and manipulated using the accompanying NEXTENGINE software, SCAN STUDIO HD version 2.0.2 (<http://www.nextengine.com/>). Machine calibration, scanning, and editing settings were based upon the recommendations (help files and tutorials) available from the user-accessed NextWiki support center, but were optimized for this particular bone and the conditions of our morphometrics laboratory. A comprehensive protocol for use on most skeletonized remains (i.e., incorporating age features on the cranium, ribs, and clavicles), with different surface treatments and degrees of completeness (e.g., smooth, porous, fractured), and under different scanning environments, is in development based on previous work by Algee-Hewitt. These guidelines will be included in the documentation for the free software package that will be released upon completion of this multipart project. Figure 2 shows representative symphyseal scans spanning the age range in the data set used here for method development. Figure 3 shows the subsequent manipulation of a scan.

New Data: The new scans were retrieved from males and females representing part of modern and archaeological populations of North America, Japan, and Thailand, Mexico, and Puerto Rico.

In the U.S., Kim spent 18 weeks at University of Tennessee collecting data from the William M. Bass Donated Collection. Upon completing data collection at UTK, she visited University of New Mexico for three weeks and collected scans from the collection of the Maxwell Museum. Kim traveled to Japan and made scans of data from the collections of the Kyoto University and Jikei University. In Thailand, data were obtained from Chiang Mai and Khon Kaen Universities. Taking advantage of contacts made in the course of this work, Kim visited and collected data from the Ministry of National Defense Agency for KIA (Killed in Action) Recovery and Identification (MAKRI), Seoul, South Korea representing young Korean soldiers killed during the Korean War.

In Mexico, Figueroa Soto spent 2 weeks at the *Universidad Nacional Autónoma de México, School of Medicine* collecting data from a skeletal collection that originates from bodies donated to the Medical Examiner's Office for the sole purpose of medical education and research. Figueroa Soto traveled for approximately three weeks to the *Instituto de Ciencias Forenses* in San Juan, Puerto Rico to collect data from a skeletal collection that originates from the body donation program at the institute.

From Windover Collection at FSU that represents one of the oldest Native American collections in the U.S., 51 new scans of the pubic symphysis were collected.

The number and types of scans collected are reported in the Results section below.

4.2. Age-At-Death Estimation Methods

Slice-Algee-Hewitt Score (SAH)

The Slice-Algee-Hewitt score (Slice and Algee-Hewitt, 2015) is based on the variability of points (vertices) orthogonal to the symphyseal surface as captured in 3D surface scans. Surface scans record the x, y, z coordinates of points on a surface as produced by a laser or structured-light scanner, or other means. Triangles (faces) are defined connecting the points so that filling in the triangle gives the appearance of the original surface when rendered on a computer screen or printer (Figure 2). The actual data are the underlying vertex coordinates (Figure 3b).

A principal components analysis of the vertices using the coordinates as variables produces an alignment such that the first principal component aligns with the greatest direction of variation in the coordinates, the second with the second greatest, and the remainder with the third (Figure 3c,d). Owing to the overall shape of the pubic symphysis, this aligns the antero-posterior direction of the symphysis with the first principal component, the dorso-ventral direction with the second principal component. The third represents variability orthogonal to the first two, which in this case coincides with surface variability. The eigenvalue, a measure of variance, associated with this third principal component is the SAH score. Slice and Algee-Hewitt (2015) showed this score to be an objective and reasonable predictor of known age: p-value = 0.0147, R-squared = 0.1432, RMSE = 17.15 years (Figure 4). Enhanced samples in later publications improved this result.

Bending Energy (BE)

The Bending Energy measure utilizes the thin-plate splines algorithm that models the bending of an infinitely flat plate to match another surface and measures the minimum energy required for the transformation (Bookstein, 1989). It was introduced as an alternative to the SAH-Score due to the fact that it takes into account the special structure of the data. For instance, the two structures illustrated in Figure 5 would produce the same SAH-Scores but the left one will be associated with a higher Bending Energy value. The implementation of the algorithm requires standardizing the size and position of

the scans. The scans are standardized to be of uniform length and the width and depth for each scan are scaled accordingly to preserve the aspect ratio. The scans are also rotated by the use of Principal Component Analysis (PCA) so that the length, width and depth of the scans are aligned along the x, y and z axes respectively. Unlike the SAH-Score, which uses all surface vertices, the BE algorithm requires the selection of equidistant vertices on the scan surface as shown in Figure 6. Similar to the SAH-Score high BE values are associated with younger individuals for whom the symphyseal surface is covered by deep furrows and ridges and lower BE values are typical for older individuals. Stoyanova et al. (2015) first introduced the BE measure in a study based on 44 scans from White males. The results showed that the measure is associated with the age-at-death (Figure 7) and reported a p-value = 0.0002, R-squared = 0.2270, and RMSE = 18.6183 years.

Ventral Curvature (VC)

Both the SAH-Score and BE capture age-related shape changes that occur on the symphyseal surface. In addition, changes also occur on the outline of the pubic symphysis. Stoyanova et al. (2017) introduced a new measure that focused on the ventral margin of the pubic symphysis. This new shape score, Ventral Curvature, measures the curvature of the ventral margin. The study shows that lower curvature values are associated with younger ages while more curved margins are observed in older individuals (Figure 8). Overall, the VC measure does not outperform either of the SAH or BE scores (Figure 9) but when combined in multiple regression models with the surface measures it contributed to the accuracy of age estimation.

5. RESULTS

5.1. Statement of the Results

New Data: The new scans were retrieved from males and females representing part of modern and archaeological populations of North America, Japan, and Thailand.

From the William M. Bass Donated Collection at the University of Tennessee, Knoxville, 273 new scans, including 146 pubic symphysis scans, 49 auricular surface scans, 48 clavicle scans, and 30 rib scans were collected. From the University of New Mexico 215 new scans, including 178 pubic symphysis scans, 19 auricular surface scans, 13 clavicle scans, and five rib scans were collected. In Japan, 200 new scans of the pubic symphysis were collected from Kyoto University and 153 new scans from the pubic symphysis from Jikei University.

In Thailand, 145 new scans, including 135 pubic symphysis scans and 10 auricular surface scans were collected from Chiang Mai University and 193 new scans, including 179 pubic symphysis scans and 14 auricular surface scans, were collected from Khon Kaen University.

From Korea, 24 new scans, including 10 pubic symphysis scans and 14 auricular surface scans from the MAKRI Collection representing young Korean soldiers killed during the Korean War were collected.

From Mexico, 43 new scans of the pubic symphysis (8 females, 20 males) were collected at the *Universidad Nacional Autónoma de Mexico, School of Medicine*. From the *Instituto de Ciencias Forenses* in San Juan, Puerto Rico, 137 new scans of the pubic symphysis (8 females, 61 males) were collected.

From Windover Collection at FSU that represents one of the oldest native American collections in the U.S., 51 new scans of the pubic symphysis were collected.

The number, type, and source of new data collected for the database of laser scans collected in the course of this project are summarized in Table 1 below.

Method Comparison

The three shape measures developed by our lab capture unique age-related characteristic on the shape of the pubic symphysis. The SAH-Score and BE both capture the transitioning of the surface from being covered by furrows and ridges to becoming more flat with age. The VC measure captures the transition of the ventral margin to a more curved shape. All three measures are associated with the age of the individual. The models presented in Stoyanova et al. (2017) report p-value = 1.6×10^{-14} , R-squared value = 0.4786, and RMSE = 14.1511 years for the SAH-Score; p-value = 4.7×10^{-9} , R-squared value = 0.3151, and RMSE = 16.3831 years for the BE; and p-value = 7.6×10^{-7} , R-squared value = 0.2365 and RMSE = 16.5457 for the VC measure. Since the SAH-Score and the BE both measure the surface of the scan they capture some of the same information. However, the VC measure captures unique outline features and therefore it can be combined with each of the surface measures for an improved multi-variate model. The results from the multivariate regression models are p-value = 6.0×10^{-15} , adjusted R-squared = 0.5178, and RMSE = 13.6830 years when VC is combined with SAH; and p-value = 1.3×10^{-11} , adjusted R-squared = 0.4267, and RMSE = 15.0704 years when combined with the BE.

Software

We have developed an open-source, multi-platform program, *forAge*, that allows scientist to use our age-estimation methods. The program takes a PLY file as input from the user. It then implements all pre-processing steps including the scaling and PCA rotation of the scans. It generates the grids used by the BE and VC methods and calculates the three shape measure, SAH, BE and VC. The program also calculates 5 age-estimates for each scan based on the models proposed in Stoyanova et al. (2017). The age-estimates are calculated from the use of the three shape measure in univariate models and two multivariate models that combine each of the SAH-Score and BE with the VC. An example of sample output from the program is shown in Figure 10. The program is available for download from <http://morphlab.sc.fsu.edu/>.

Reliability

In Stoyanova et al. (2015), we showed the increased repeatability of the five computational methods based on the three shape measures through a single (intra) observer error test. However, we did not fully and systematically substantiate the methods' performance and reliability at the level of multiple observers with different training backgrounds and/or different experience levels. In response to this concern, we investigated the intra-scan variability and within- and between observer reliability in initial scan data capturing and editing using 3D laser scans of the Suchey-Brooks pubic symphysis casts. For this study, the upper and lower stages of each of the six Suchey-Brooks phases were scanned three times by a single observer (n = 36).

Five observers with various training background and experience levels independently edited the triplicate set of scans over three consecutive trials. Of the five, three are biological anthropologists and represent a senior researcher/the developer of the scan editing protocol (Observer 1), a newly-graduated Ph.D. (Observer 3), and an advanced graduate student/forensic investigator (Observer 4). All three had received training in skeletal biology/age estimation and have practiced the computational methods prior to this study. Observer 2 is a computational scientist and the developer of the computational methods and accompanying software, *forAge*, with some knowledge in osteology and pubic symphyseal scan editing experience. Later in the study, the fifth observer, an undergraduate student with limited knowledge in age estimation using skeletal remains, was introduced. Observer 5 was blinded to the study's purpose and provided with only minimal instruction on scan editing via online texts. From each of the edited scans, shape measures and final age estimates were obtained and these values were subjected to a calculation of intraclass

correlation coefficients (ICC) which was employed as a measure of observer agreement among various scans.

In addition, we simulated two extra editing situations where the practitioner misidentifies age-related traits due to unfamiliarity with the scan editing protocol and/or inexperience with skeletal biology. These conditions specifically address following two questions regarding scan editing; (i) how far one should delete the margin around the symphyseal face for the scan to be useful for age estimation (i.e. 2mm vs. 4mm vs. 1cm); and (ii) whether one should include or exclude the pubic tubercle, an age-informative trait that sometimes extends to the symphyseal face making it difficult to identify the exact border when editing raw scans. As for the second concern, we hypothesized that either inclusion or exclusion of the pubic tubercle will influence the VC value as the feature protrudes ventrally. The pairwise t-test was conducted to evaluate the mean difference in age estimates derived from properly edited scans vs. those improperly edited on purpose.

Our results show that (i) five observers with various training background and experience levels edited the scans consistently for all three trials and the derived shape measures and age estimates were in excellent agreement among observers ($0.76 < ICC < 0.86$ without Observer 5, $0.60 < ICC > 0.75$ with Observer 5), and (ii) the computational methods are robust to a measured degree of scan trimming error. This more complex observer error study not only corroborates the results of the single observer error study presented in Stoyanova et al. (2015), but also supports the application of computational methods to 3D laser scanned images for reliable age-at-death estimation, with reduced subjectivity.

Asymmetry

Each person has two pubic symphyseal surfaces, left and right, that follow similar patterns of aging but are not identical. Within-individual asymmetry is an important factor in age-at-death estimation. Traditional techniques put emphasis on studying left surfaces but it is important to evaluate the effect of asymmetry on our computational methods. We performed a study based on 88 White males for whom both the left and right symphyseal surfaces were scanned. In Stoyanova et al. (2018), we present results that show that the computational algorithms are not sensitive to the within-individual asymmetry. Neither side produced consistently better (or worse) results. Each individual's level of asymmetry was defined as the difference in years between the left and right estimates. The asymmetry values were then regressed against the

individuals' weight, height, body-mass index and advanced age as a way to explain the higher asymmetry observed for some individuals. None of the four factors was associated with the level of asymmetry. Our findings indicate that scientists can with the same level of confidence generate an age-estimate using either of the left or right sides which is useful in situations when one side is missing or damaged. Further, the lack of sensitivity of the methods to individuals' life-history factors would allow investigators to rely on the age-estimates when there is no other known information about the person in question.

6. CONCLUSION

6.1. Discussion of Findings

This project, MODELING SURFACE MORPHOLOGY OF THE PUBIC SYMPHYSIS, has sought to advance both the theory and method of Forensic Anthropology by developing objective, fully-computational methods for the estimation of age-at death from the adult skeleton. Indicators of age that are known to undergo degenerative changes (as bone build up and break down) with advancing time, are widely used to estimate chronological age-at-death from skeletal remains recovered in medico-legal contexts. In conventional practice, involving the gross morphological assessment of the bone feature or surface, the indicator is visually compared against some set of criteria, a corresponding score or phase is assigned, and age-at-death is estimated from the age-range associated with the chosen phase. While appealing for its ease of application – as it permits straightforward explanation, uses limited laboratory equipment (casts or images or text descriptions of exemplar phases), and requires no mathematical computations, this morphoscopic process of bone-to-phase matching is prone to methodological subjectivity and disagreement among observers, which together introduce estimation error. We have argued that there is both great need and untapped potential for developing alternative tools for the quantification of age-related change in skeletal morphology that (1) better approximate true chronological age, regardless of the individual case characteristics and the nature of the forensic practitioner's experience, and (2) better meet current expectations for standards of evidence in the medico-legal casework context.

We have proposed, therefore, a novel approach to age estimation that uses laser-scan-derived, three-dimensional coordinate data to quantify the shape of the skeletal element or its feature(s), in order to capture the senescent changes in morphology occurring over the adult lifespan with reduced error, high

precision, accuracy and repeatability, and low requirements for training or experience. We have developed a comprehensive framework for age indicator analysis and the prediction of age that provides robust estimates of chronological age by bringing the morphometric, mathematical, and programming strengths of scientific computing to bear on the skeletal indicators routinely used in the macromorphoscopic determination of age. Building upon the success of our preliminary work (Slice and Algee-Hewitt, 2015), we have developed new and refined our prior methodologies for inferring age from the pubic symphysis, and we have expanded our coverage of variation in this skeletal element, including both sexes, sampling from the full range of adult ages, and applying our techniques on data we collected on many different biosocial and biogeographic populations. Specifically, we have implemented three shape algorithms that produce novel age-related measures useful for regression-based estimation. We have delivered (1) a revised version of the SAH-score method that captures the variance on the symphyseal face to quantify the gradual flattening of the surface associated with aging (Slice & Algee-Hewitt, 2015), (2) a Bending Energy (BE) method that uses thin-plate splines techniques to determine the bending energy required for transforming a perfectly flat, infinitely thin plate to match the surface of a pubic symphysis scan in order to quantify surface variance (Stoyanova et al., 2015), and (3) a Ventral Curvature (VC) method that quantifies the progressive formation of a rim around the entire symphyseal surface and its later erosion (Stoyanova et al., 2017). We have collected laser scan data on multiple age indicators from \approx 1200 modern skeletons of reported American Black, Hispanic, and White, Mexican, Puerto Rican, Spanish, Japanese, Korean, and Thai identity. These data were used to build the foundation for our computational framework by, first, setting standards for shape analysis and, second, to build regression models for age prediction from the shape measures. To these goals, we have completed work on, for example, (a) grid density and surface sampling resolution, (b) indicator partitioning schemes for fragmentary remains, (c) outline, 2D Fourier, wavelet parameterization, regression and Bayesian analyses, (d) the effects of sample sex, population diversity, asymmetry, and life-history factors on estimation, (f) rates and patterns of observer error, (g) application of our methods to the auricular surface as an alternative pelvic age indicator, (h) comparison of our results against those derived by traditional skeletal aging methods and the more recent technique of transition analysis, and (7) released free, downloadable, stand-alone software, *forAge*, that

facilitates our methods and permits the addition of new scans and age distribution data to build databases of reference samples and age priors.

We have demonstrated how it is possible to overcome the high subjectivity and error associated with traditional approaches to adult age-at-death estimation from the skeleton – typically, the macroscopic assessment of the pubic symphysis, the association of this indicator with a phase or score for which an age point estimate/ internal has been predefined – by developing a full computational, semi-automated framework that enables the objective and mathematically robust assessment of true chronological age in forensic anthropological case contexts. By focusing on the changes observed in the shape of the symphyseal surface and the curvature of ventral margin of the pubic symphysis, we have shown how we can build multivariate regression models (with root-mean-square-errors between 13.7 and 16.5 years) that produce objective, accurate and precise age-at-death estimates (Stoyanova et al, 2017). These estimates are just as good as, if not better than, results obtained with the conventional gold-standard methods, while they have the special advantage of being grounded in accepted mathematical theory and robust statistical practice. Further, we have established how our methodologies have several features desirable to forensic casework and important for meeting evolving evidentiary standards: they (1) permit quantifiable error at the level of the technique and observer, (2) allow for the more reliable application of an age estimation technique when the practitioner’s expertise is low, (3) offer straightforward repeatability and validation testing given their scan and algorithmic-driven design, and (4) provide the ability to incorporate new samples and perform independent analyses using the open-source, method-specific software, *forAge*.

In order to advocate for the reliable use of this project’s framework for age estimation in actual forensic applications, we expanded upon the single observer error test reported in Stoyanova et al (2015). In this study, we sought to quantify the possible error that may occur during the initial process of scan-data capturing and scan editing, especially considering these effects for multiple observers with various experience levels and training backgrounds. We demonstrated that observer error is low: the scans are consistently edited and highly congruent, reliable, and accurate with respect to age estimates produced using a set of test scans and applying our computational methods via the software package, *forAge*. These results show that the computational methods and the use of scan data can alleviate the concern of subjectivity that

is known to be associated with the visual assessment of age-related skeletal features (Kim et al, 2018). As age alone does not account for all the variation observed in the pubic symphysis, and presumably other skeletal indicators used in aging, and individual factors (like sex, population, and lifestyle) may affect the rate at which the pubic symphysis ages and, in turn, its morphological expression, we sought to address some of the confounding factors to estimation associated with the unique characteristics of a given case and not the technical application of the method. We pursued, therefore, studies of sex and population differences as well as tests of asymmetry. Our preliminary studies of sex and populations suggested that while pooled sample analyses using mismatched reference samples yield reasonably good estimates, even better results can be obtained when group-specific equations are built, differentiating the analysis of, for example, males and females, and Puerto Rican and American White (Figueroa Soto, forthcoming – human biology paper). Our asymmetry analysis indicated that this project’s computational, shape-based techniques are largely robust to bilateral deviations in shape. For those cases with high asymmetric pubic symphyses, we found that the differences in the final age estimates produced with our methods were not associated with such factors as advanced age, weight or stature. These results are important for forensic casework practice as they confirm for this bilateral indicator how our computational framework can be equally applied to either side (left or right) in unknown case contexts, when no life-history information is available (Stoyanova et al, 2018).

6.2. Discussion of Implications for Policy and Practice

We have provided a novel semi-automated toolkit that does not depend fully on the user, but is instead computationally driven, which allows for the direct quantification of error and the probabilistic estimation of age. Our methods are transformative to how age-at-death estimation can be approached in forensic casework contexts by demonstrating the potential of such a fully-computational framework to producing highly reliable individual estimates of age that meet medico-legal standards of evidence. At the same time this research makes significant contributions to understating age-related changes in skeletal indicator morphology across different peoples who vary widely in their life-histories, given the kind and breadth of data exploited and diversity in the samples used in our analyses. In addition to advancing new methodological system, we have also contributed to the field a wealth of new information available for future research. Our project data consists of the raw laser scans of several skeletal elements (including the pubic symphysis, auricular surface,

and rib ends) that are a priori known to be informative of age and, so, have served as the basis for many of the classic age assignment techniques. It also includes the 3D (x,y,z) coordinates extracted from these laser scans. We have also produced the first comprehensive data set of age estimates obtained by applying the standard (pelvic, rib, clavicle) methods, transition analysis, and our fully computational approach to the same laser scanned cases. Our sample is, to our knowledge, unprecedented, given its size and biogeographic diversity: while we sought to cover the breadth of variation in the major American groups, given the emphasis on forensic casework standards and methods in the United States, we also prioritized data collection on Latino and Asian populations, recognizing that their patterns and rates of senescence are poorly known, they are the largest and fastest growing minority groups in the U.S., and they represent the majority of the undocumented [im]migrants in forensic casework.

Although scanners are increasingly affordable, easily run, and widely used in many fields of biological sciences, including anthropology, they have received limited attention for bone surface modeling and no such data has been previously acquired for age-estimation. We believe that our study has demonstrated the merit of embracing these technological advances as new standards and, in turn, the potential of effecting positive change to conventional forensic anthropological protocol. Our work can serve as justification for not only incorporating laser-scan-based, semi-automated and fully computational methodologies into forensic practice, but also for demanding a new baseline for standards of evidence. Specifically, it can inform procedures for documenting skeletal indicators of age, producing chronological age estimates, and preserving and sharing data, results, and the mathematics, algorithms and programming code that underlie the implementation of the proposed computational methods. Laser scanning has several advantages for forensic practice, as it enables accurate measurement, eliminates contact with the specimen, simplifies data sharing, permits virtual modeling and reconstruction, and offers digital solution to the long-term archiving of casework remains. The nature of the scan-derived data and the open-source design of our software (with free access to the methods included therein), gives researchers the flexibility to develop new computational shape-analysis methods. Such numerical analysis has notable value as it permits fast big data processing, codifies techniques, and improves accuracy in estimation. It also has the ability to statistically account for case characteristics that may

otherwise confound estimation just as it has the important flexibility to allow for its tailoring to different forensic contexts.

6.3. Discussion of Future Work

This project has laid the foundation for various lines of future research. First, our preliminary testing suggests that there is great promise in continuing to refine our current computational shape algorithms. We will, therefore, test various modifications to our framework on the skeletal samples that we have already collected, with a focus on the understudied Asian populations. Second, we will expand the scope of variables tested for life history effects and investigate the benefit of including any of factors as covariates in our prediction models. Third, we will use our currently available pubic symphysis data to assess in greater depth the accuracy of age estimation from fragmentary remains and develop a computational solution for handling missing scan data. To achieve this goal, we will artificially generate fragmentary scans by deleting sections of the full scans. We will aim to assess the percentage of missing information allowed, as well as any particular areas that need to be present, for reliable and accurate age estimation. We will then develop algorithms that either impute the coordinate data for the missing regions, thereby, reconstructing the element, or estimate age in the presence of these lacunae. Fourth, we will investigate the applicability of our current algorithms with other skeletal indicators of adult age, distributed throughout the entire skeleton. By extension, we will generate new predictive models for these additional indicators, just as we will develop new regression equations for other sampled populations. Fifth, we intend to expand upon our current reference sample by building a free-access, online database of scan data that covers the range of spatial, temporal and bio-social group variation that we might expect to encounter in forensic casework at present and in the future. Finally, we plan to test the applicability of machine learning methods, like deep neural networks, for improved inference (Chen et al, 2016; Liang et al, 2017; Lim et al, 2016; Omodaka et al, 2018; Wade et al, 2017). We posit that, in finding data correlations that evade human intuition, we can produce models that give age-range and point-estimate predictions superior to any of the classic statistical methods. This machine learning approach has, therefore, the potential to further decrease the estimation error in age estimation and to better capture more nuanced morphological differences across a wide variety of skeletal indicators. Recent publications implementing this analytical approach have shown encouraging results for skeletal age estimation, albeit not using laser-scan data or adopting a fully-computational framework of analysis (Buk

et al, 2012; Corsini et al, 2005; Navega et al, 2017). However, preliminary research by Algee-Hewitt and Kim (presented at the Wenner-Gren Workshop) has demonstrated, with neural networks how the laser scan and our shape algorithm data obtained for our age estimation project can be used to infer population of origin, as proportions of continental ancestry. This novel work supports not only the value of investigating this and other AI-ML approaches but the possibility of revealing other information on personal identity from coordinate data on the pubic symphysis as well as other classic skeletal features.

6.4. Tables

COLLECTION	LOCATION	pubic symphysis (ps)	Ind. (ps)	auricular surface (as)	Ind. (as)	clavicular end (ce)	Ind. (ce)	sternal 4th rib end (s4r)	Ind. (s4r)	SCAN TOTAL
William M. Bass Donated Collection	UTK	146	73	49	26	48	26	30	17	273
Maxwell Museum Collection	University of New Mexico	178	95	19	10	13	7	5	3	215
Windover	Tallahassee, FL	51	0	0	0	0	0	0	0	51
Kyoto University	Khoto, Japan	200	136	0	0	0	0	0	0	200
Jikei University	Tokyo, Japan	153	151	0	0	0	0	0	0	153
MAKRI	Seoul, South Korea	10	9	14	14	0	0	0	0	24
Chiang Mai University	Chiang Mai, Thailand	135	135	10	10	0	0	0	0	145
Khon Kaen University	Khon Kaen, Thailand	179	169	14	14	0	0	0	0	193
Universidad Nacional Autónoma de Mexico	Mexico City, Mexico	43	21	0	0	0	0	0	0	43
Instituto de Ciencias Forenses	San Juan, Puerto Rico	137	69	0	0	0	0	0	0	137
SCAN TOTAL		1232	858	106	74	61	33	35	20	1281

6.5. Figures

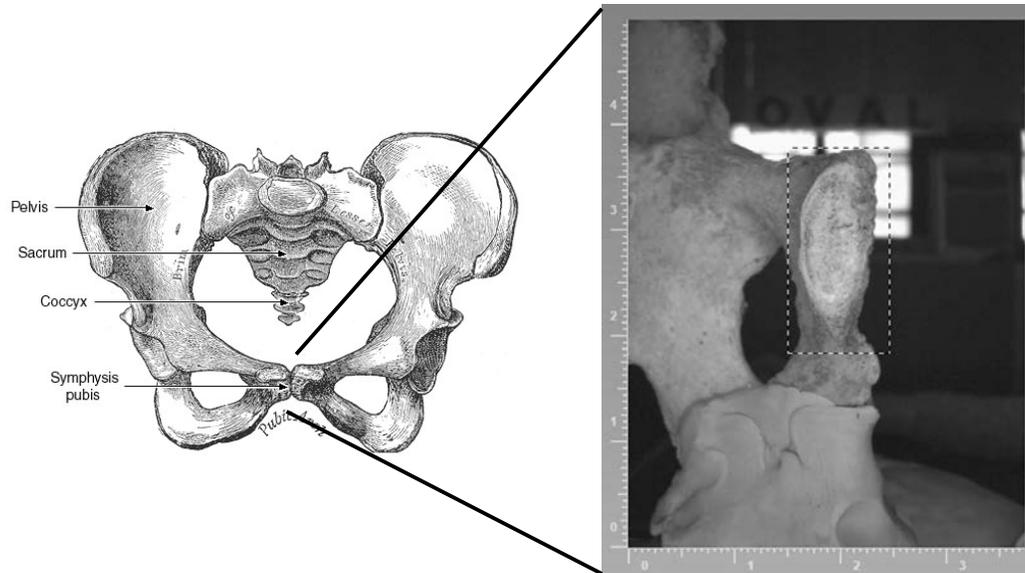


Figure 1: Pubic symphysis anatomical location (left) and mounted for scanning (right). Drawing (left) from https://commons.wikimedia.org/wiki/File:Symphysis_Pubis.png.

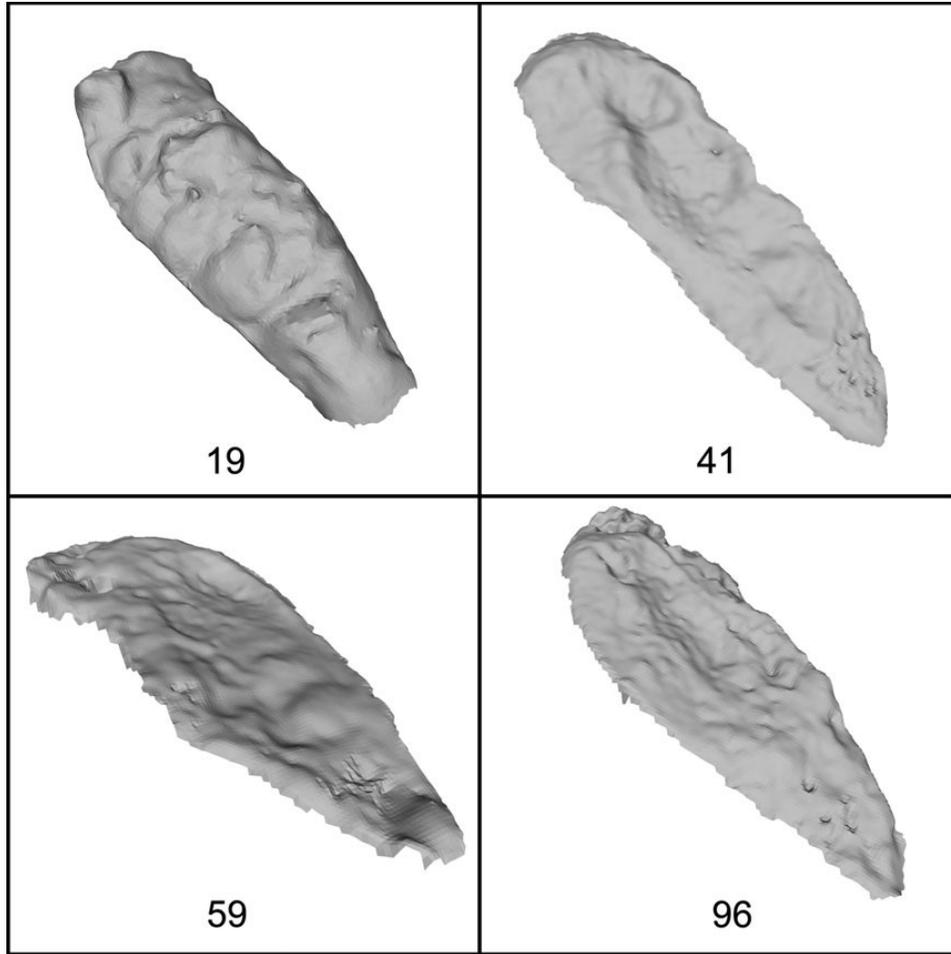


Figure 2: Age-related variation in symphyseal surface. Age in years shown below scan.

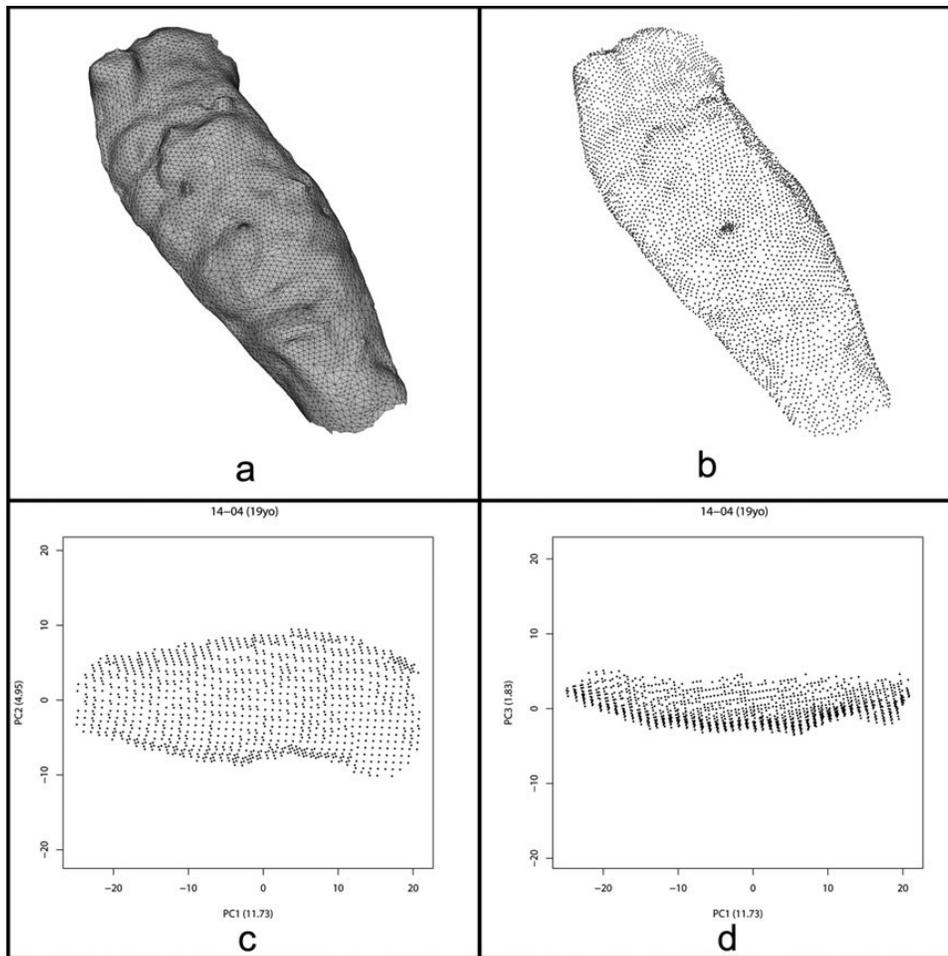


Figure 3: When rendered (see Figure 2) surfaces appear smooth. They are, in fact, shaded triangles formed by vertices (a). The vertices of these triangles (b) are the raw data obtained by the scanner and subject of analytical methods for age determination. For the SAH method (see text) the vertex coordinates are subject to a principle components analysis. The first two PCs (c) correspond roughly to the superior-inferior axis (PC1) of the surface and the dorso-ventral axis (PC2). The third PC (d) encodes variation in the symphyseal surface exploited by the SAH method.

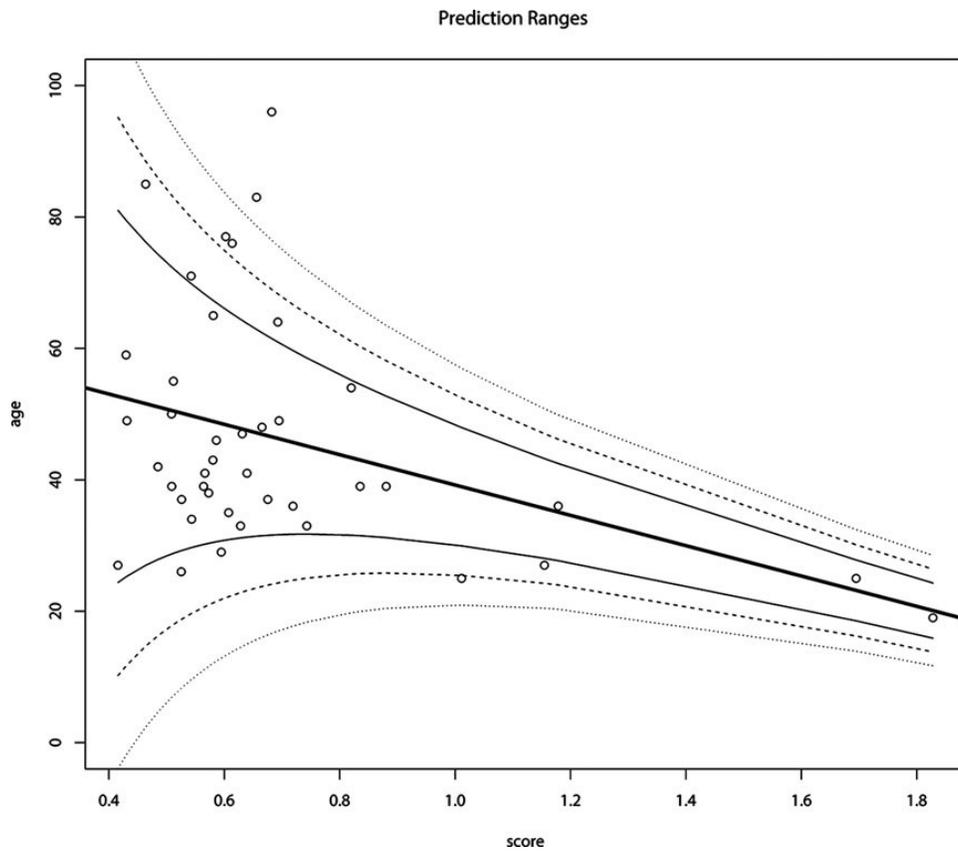


Figure 4: Association between SAH score (x-axis) and age (y-axis). See text for details. Figure from Slice and Algee-Hewitt (2015).

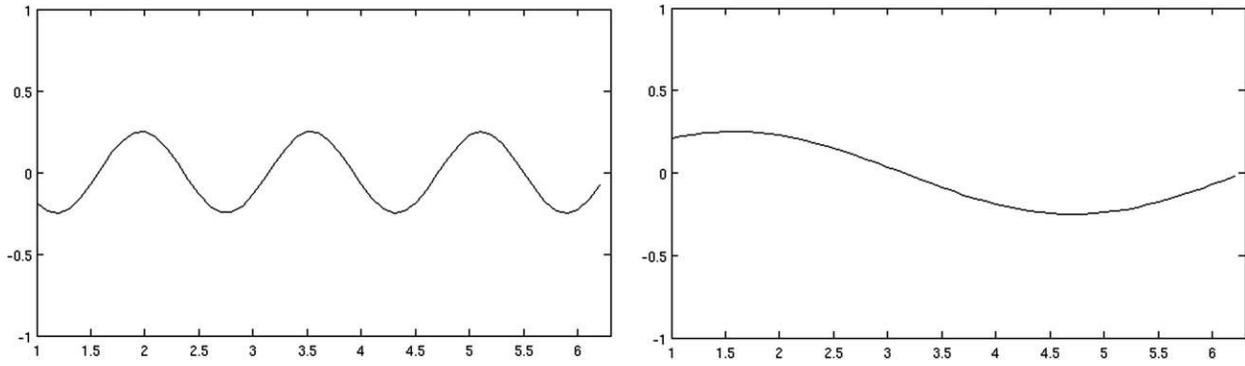


Figure 5: Different surface variation leading to identical SAH scores. See text for details. Figure from Stoyanova et al. (2015).

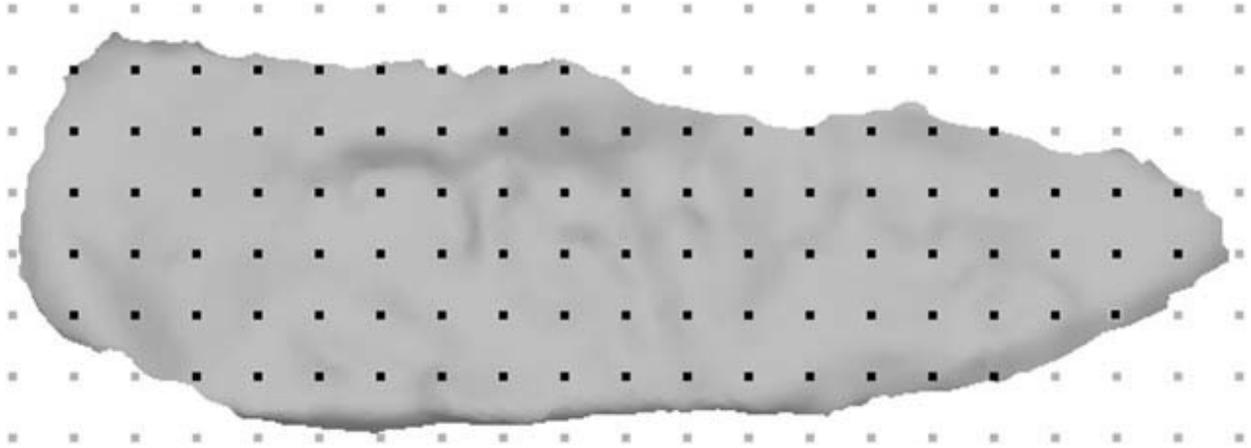


Figure 6: Regular grid generated for symphyseal surface aligned to spatial principle components. Interior points (black) are projected onto the surface. The parameters for the distortion of a thin-plate in the x - y plane to match the heights of these points is computed and the net bending energy of that distortion used as a measure for age-association analysis. Figure from Stoyanova et al. (2015).

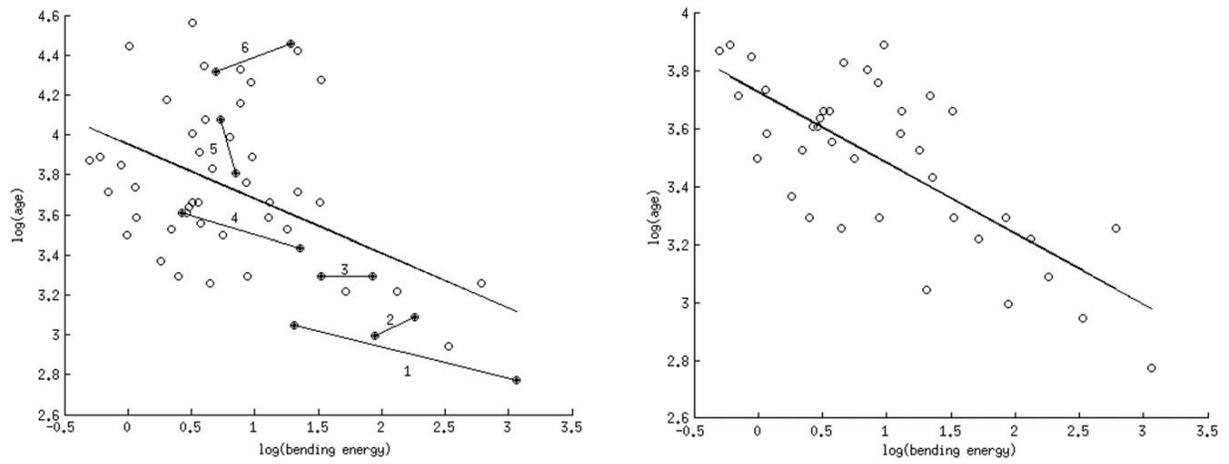


Figure 7: Association of $\log(\text{age})$ and $\log(\text{BE})$ from the Thin-Plate Spline/Bending Energy method. Left shows association for all data. Lines connect pairs of Suchey-Brooks surface scans. Right shows association when age is restricted to <50 . Figure from Stoyanova et al. (2015).



Figure 8: Association of age with ventral curvature (VC) exploited for age-at-death prediction. Ages represented (left to right): 25, 36, 85. Figure from Stoyanova et al. (2017).

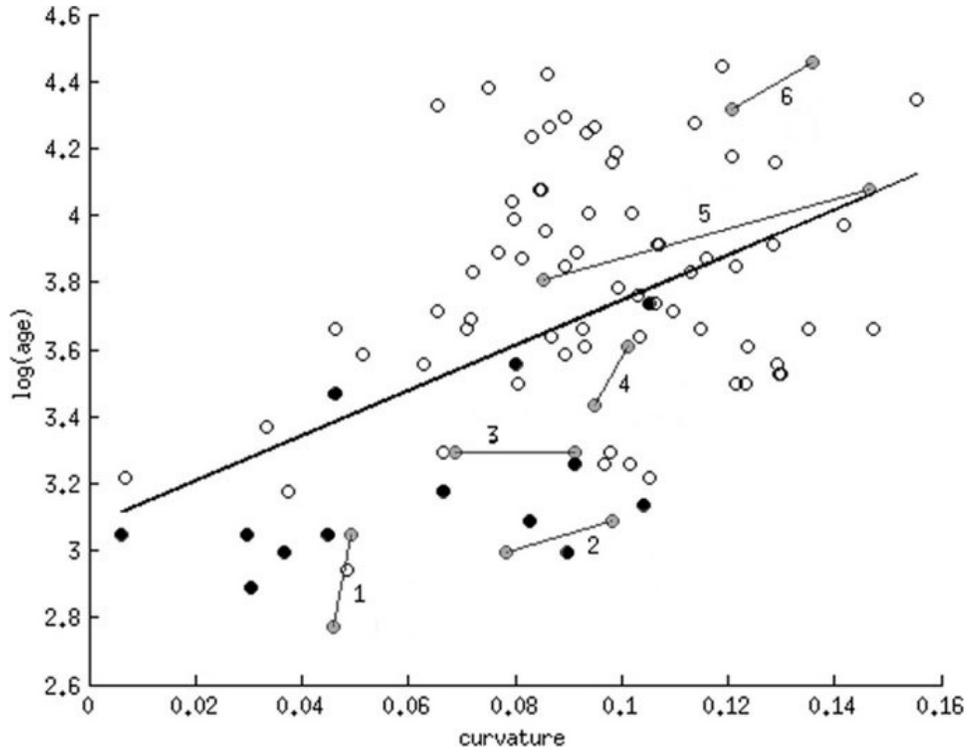


Figure 9: Association of $\log(\text{age})$ with $\log(\text{VC})$. Gray symbols show connected pairs of Suchey-Brooks casts - phase indicated above connecting line. Black symbols show McKern and Stewart casts. Figure from Stoyanova et al. (2017).

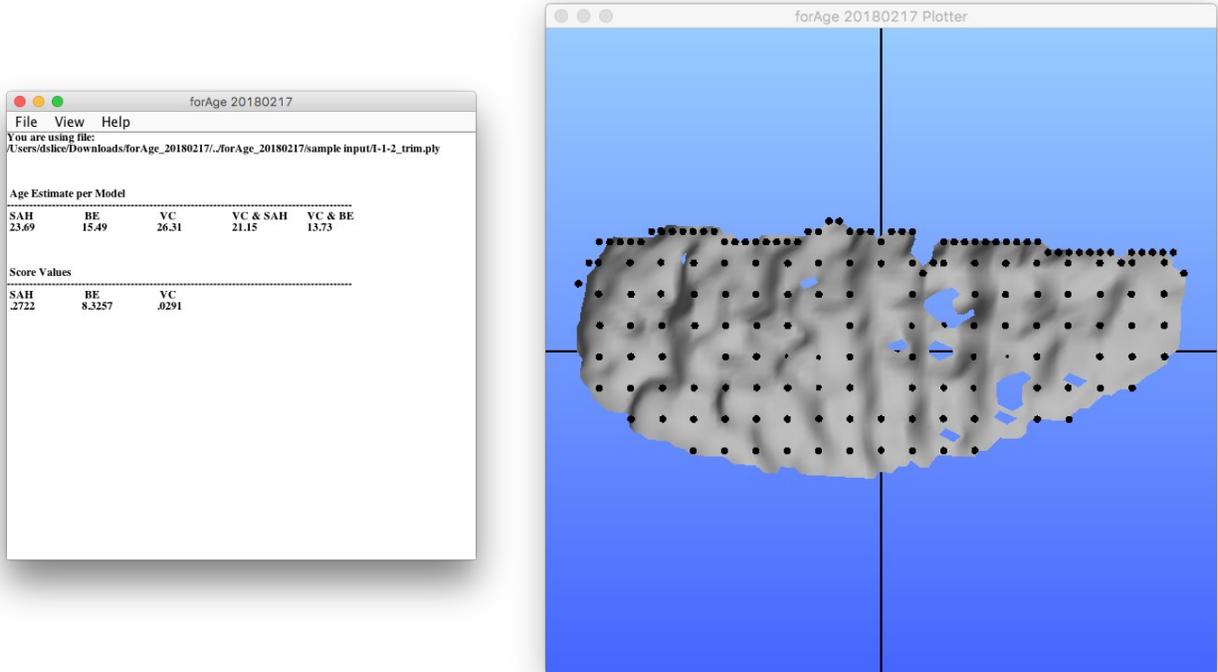


Figure 10: forAge: software for age-at-death estimation implementing methods developed in the course of this research. Left: output window showing estimates for SAH, BE, VC, and multiple regression of age onto SAH+VC and BE+VC. Right: display window showing symphyseal surface, BE grid points, and VC edge points (dense points along top).

7. REFERENCES

Boldsen JL, Milner GR, Konigsberg LW, Wood JW. 2002. Transition Analysis: A new method for estimating age from skeletons. In: Hoppa RD, and Vaupel JW, editors. *Paleodemography: Age distributions from skeletal samples*. Cambridge, UK: Cambridge University Press.

Adams BJ, and Byrd JE. 2008. *Recovery, analysis, and identification of commingled human remains*: Springer Science & Business Media.

Algee-Hewitt BFB. 2011. "If and How Many 'Races'? The Application of Mixture Modeling to World-Wide Human Craniometric Variation." PhD diss., University of Tennessee. http://trace.tennessee.edu/utk_graddiss/1165

Algee-Hewitt BFB. 2017a. Geographic substructure in craniometric estimates of admixture for contemporary American populations. *American Journal of Physical Anthropology*, 164(2):260-280.

Algee-Hewitt BFB. 2017b. Temporal trends in craniometric estimates of admixture for a modern American sample. *American Journal of Physical Anthropology*, 163(4):729-740.

Algee-Hewitt BFB, Hughes CE, and Anderson BE. 2018. Temporal, Geographic and Identification Trends in Craniometric Estimates of Ancestry for Persons of Latin American Origin. *Forensic Anthropology*, 1(1):4-17.

Bookstein, F. 1989. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6) , 567-585.

Brooks S, and Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, 5(3):227-238.

Buckberry JL, and Chamberlain AT. 2002. Age Estimation From the Auricular Surface of the Ilium: A Revised Method. *American Journal of Physical Anthropology*, 119:231-239.

Buk Z, Kordik P, Bruzek J, Schmitt A, and Snorek M. 2012. The age at death assessment in a multi-ethnic sample of pelvic bones using nature-inspired data mining methods. *Forensic Science International*, 220(1-3):294. e291-294. E299.

Chen D, Sarkar S, Candia J, Florczyk SJ, Bodhak S, Driscoll MK, ... Losert W. 2016. Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues. *Biomaterials*, 104:104-118.

Corsini M-M, Schmitt A, and Bruzek J. 2005. Aging process variability on the human skeleton: artificial network as an appropriate tool for age at death assessment. *Forensic Science International*, 148(2-3):163-167.

DiGangi EA, Bethard JD, Kimmerle EH, and Konigsberg LW. 2009. A new method for estimating age-at-death from the first rib. *American Journal of Physical Anthropology*, 138(2):164-176.

Djurić M, Djonić D, Nikolić S, Popović D, and Marinković J. 2007a. Evaluation of the Suchey–Brooks method for aging skeletons in the Balkans. *Journal of Forensic Sciences*, 52(1):21-23.

Djurić M, Djonić D, Djonic D, and Skinner M. 2007b. Identification of victims from two mass-graves in Serbia: a critical evaluation of classical markers of identity. *Forensic Science International*, 172(2-3):125-129.

Falys CG, Schutkowski H, and Weston DA. 2006. Auricular surface aging: worse than expected? A test of the revised method on a documented historic skeletal assemblage. *American Journal of Physical Anthropology*, 130(4):508-513.

Fojas CL, Kim J, Minsky-Rowland JD, and Algee-Hewitt B. 2015. Testing Inter-observer Reliability of the Transition Analysis Aging Method on the William M. Bass Forensic Skeletal Collection. 67th annual meeting of the American Academy of Forensic Sciences. Orlando, Florid. P 118.

Frey, W. 2015. Front Matter. In *Diversity Explosion: How New Racial Demographics are Remaking America* (pp. I-VI). Brookings Institution Press.

Garvin HM, and Passalacqua NV. 2012. Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *Journal of Forensic Sciences*, 57(2):427-433.

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. *Journal of Forensic Sciences*, 55(5):1152-1156.

Hens SM, and Belcastro MG. 2012. Auricular surface aging: A blind test of the revised method on historic Italians from Sardinia. *Forensic Science International*, 214(1-3):209.e201-209.e205.

Hens SM, Rastelli E, and Belcastro G. 2008. Age estimation from the human os coxa: a test on a documented Italian collection. *Journal of Forensic Sciences*, 53(5):1040-1043.

Hoppa RD. 2000. Population variation in osteological aging criteria: An example from the pubic symphysis. *American Journal of Physical Anthropology*, 111(2):185-191.

Hughes CE, Algee-Hewitt BF, Reineke R, Clausing E, and Anderson BE. 2017. Temporal patterns of Mexican migrant genetic ancestry: Implications for identification. *American Anthropologist*, 119(2):193-208.

Igarashi Y, Uesu K, T. W, and E. K. 2005. New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology*, 128:324-339.

Işcan. 1987. Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences*, 32:452-466.

Işcan MY, Loth SR, and Wright RK. 1984. Age estimation from the rib by phase analysis: white males. *Journal of Forensic Sciences*, 29(4):1094-1104.

Işcan MY, Loth SR, and Wright RK. 1985. Age estimation from the rib by phase analysis: white females. *Journal of Forensic Sciences*, 30(3):853.

Jantz LM, and Jantz RL. 1999. Secular change in long bone length and proportion in the United States, 1800–1970. *American Journal of Physical Anthropology*, 110(1):57-67.

Kemkes-Grottenthaler A. 2002. Aging through the ages: Historical perspectives on age indicator methods. In: Hoppa RD, Vaupel, J. W. , editor. *Paleodemography: Age distributions from skeletal samples*. Cambridge.

Kim J, Algee-Hewitt, BFB, Stoyanova DK, Figueroa-Soto C, and Slice DE. 2018. “Testing Reliability of the Computational Age-At-Death Estimation Methods between Five Observers Using Three-Dimensional Image Data of the Pubic Symphysis,.” *Journal of Forensic Sciences* 0, no. 0. <https://doi.org/10.1111/1556-4029.13842>.

Kimmerle EH, Konigsberg LW, Jantz RL, and Baraybar JP. 2008a. Analysis of Age-at-Death Estimation Through the Use of Pubic Symphyseal Data. *Journal of Forensic Sciences*, 53(3):558-568.

Kimmerle EH, Prince DA, and Berg GE. 2008b. Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth. *Journal of Forensic Sciences*, 53(3):594-600.

- Klales AR. 2016. Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Sciences*, 61(2):295-301.
- Klepinger L. 2001. Stature, maturation variation and secular trends in forensic anthropology. *Journal of Forensic Sciences*, 46(4):788-790.
- Komar DA, and Grivas C. 2008. Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. *American Journal of Physical Anthropology*, 137(2):224-233.
- Konigsberg LW. 2015. Multivariate cumulative probit for age estimation using ordinal categorical data. *Annals of Human Biology*, 42(4):368-378.
- Konigsberg LW, Herrmann NP, Wescott DJ, and Kimmerle EH. 2008. Estimation and evidence in forensic anthropology: age-at-death. *Journal of Forensic Science*, 53(3):541-557.
- Langley-Shirley N, and Jantz RL. 2010. A Bayesian Approach to Age Estimation in Modern Americans from the Clavicle. *Journal of Forensic Sciences*, 55:571-583.
- Liang L, Liu M, Martin C, Elefteriades JA, Sun W. 2017. A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm. *Biomechanics And Modeling In Mechanobiology*, 5:1519-1533.
- Lim I, Gehre A, Kobbelt L. 2016. Identifying Style of 3D Shapes using Deep Metric Learning. *Computer Graphics Forum*, 35(5):207-215.
- Lovejoy CO, Meindl RS, Pryzbeck TR, and Mensforth RP. 1985. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, 68:15-28.
- McKern TW, and Stewart TD. 1957. Skeletal Age Changes in Young American Males: Analyzed from the Standpoint of Age Identification. *Quartermaster research and development command*.
- Meindl RS, Lovejoy CO, Mensforth RP, Walker RA. 1985. Revised Method of Age Determination Using the OS Pubis, with a Review and Tests of Other Current Methods of Pubic Symphyseal Aging. *Am J Phys Anthropol* 68: 29-45.
- Merritt CE. 2014. A test of Hartnett's revisions to the pubic symphysis and fourth rib methods on a modern sample. *Journal of Forensic Sciences*, 59(3):703-711.

Milner GR, and Boldsen JL. 2017. Life not death: Epidemiology from skeletons. *International Journal of Paleopathology*, 17:26-39.

Milner GR, Wood JW, and Boldsen JL. 2008. Advances in Paelodemography. In: Katzenberg MA, and Saunders SR, editors. *Biological Anthropology of Human Skeleton*. 2nd ed. p 561–600. New York, NY: John Wiley and Sons, Inc.

Murray KA, and Murray T. 1991. A Test of the Auricular Surface Aging Technique. *Journal of Forensic Sciences*, 36:1162-1169.

National Research Council. 2009. *Strengthening forensic science in the United States: a path forward*: National Academies Press.

Navega D, Coelho JdO, Cunha E, and Curate F. 2017. DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. *Journal of Forensic Science*.

Nawrocki S. 2010. The nature and sources of error in the estimation of age at death from the skeleton. *Age estimation of the human skeleton*,79-101.

Omodaka, K., An, G., Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., & ... Nakazawa, T. 2018. Classification of optic disc shape in glaucoma using machine learning based on quantified ocular parameters. *Plos ONE*, 12(12):1-10.

Ortner DJ. 2003. *Identification of pathological conditions in human skeletal remains*. San Diego, CA: Academic Press.

Osborne DL, Simmons TL, and Nawrocki SP. 2004. Reconsidering the auricular surface as an indicator of age at death. *Journal of Forensic Sciences*, 49(5):905-911.

Rissech C, Wilson J, Winburn AP, Turbon D, and Steadman D. 2012. A comparison of three established age estimation methods on an adult Spanish sample. *International Journal of Legal Medicine*, 126(1):145-155.

Ritz-Timme S, Cattaneo C, Collins M, Waite E, Schütz H, Kaatsch H-J, and Borrman H. 2000. Age estimation: the state of the art in relation to the specific demands of forensic practise. *International Journal of Legal Medicine*, 113(3):129-136.

San Millán M, Rissech C, and Turbón D. 2013. A test of Suchey–Brooks (pubic symphysis) and Buckberry–Chamberlain (auricular surface) methods on an identified Spanish sample: paleodemographic implications. *Journal of Archaeological Science*, 40(4):1743-1751.

Saunders SR, Fitzgerald C, Rogers T, Dudar C, and McKillop H. 1992. A test of several methods of skeletal age estimation using a documented archaeological sample. *Journal of Forensic Science*, 2:97-118.

Schmitt A. 2004. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. *International Journal of Osteoarchaeology*, 14:1-6.

Shirley NR, and Ramirez Montes PA. 2015. Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component-Based Methods. *Journal of Forensic Sciences*, 60(1):107-111.

Shrestha L. 2011. *Changing Demographic Profile of the United States*. DIANE Publishing.

Slice DE, Algee-Hewitt BFB. 2015. Modeling Bone Surface Morphology: A Fully-Quantitative Method for Age-At-Death Estimation Using The Pubic Symphysis. *Journal of Forensic Science* 60(4):835-43.

Spradley MK, Jantz RL, Robinson A, and Peccerelli F. 2008. Demographic Change and Forensic Identification: Problems in Metric Identification of Hispanic Skeletons*. *Journal of Forensic Sciences*, 53(1):21-28.

Steadman DW, Adams BJ, and Konigsberg LW. 2006. Statistical basis for positive identification in forensic anthropology. *American Journal of Physical Anthropology*, 131(1):15-26.

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. *American Journal of Physical Anthropology* 158:431-440.

Stoyanova D, 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. *Journal of Forensic Sciences* 62:1434-1444.

Stoyanova D, Algee-Hewitt BFB, Kim J, Slice DE. 2018. A study on the asymmetry of the human left and right pubic symphyseal surfaces using high-definition data capture and computational shape methods. Under revision at the *Journal of Forensic Sciences*.

Stoyanova, Detelina K., Bridget F. B. Algee-Hewitt, Jieun Kim, and Dennis E. Slice. 2018. "A Study on the Asymmetry of the Human Left and Right Pubic Symphyseal Surfaces Using High-Definition Data Capture and Computational

Shape Methods.” *Journal of Forensic Sciences* 0, no. 0.
<https://doi.org/10.1111/1556-4029.13871>.

Suchey, J.M., Katz, D. 1998. Applications of pubic age determination in a forensic setting. In: K.J. Reichs (Ed.), *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edition, C. C. Thomas, Springfield, IL. 204–236.

SWGANTH. 2013. *Scientific Working Group for Forensic Anthropology: Age Estimation* p1-5.

Taylor, P. 2014. *The next America: boomers, millennials, and the looming generational showdown*. First edition. New York: PublicAffairs.

Ubelaker DH, Owsley DW, Houck MM, Craig E, Grant W, Woltanski T, Fram R, Sandness K, and Peerwani N. 1995. The role of forensic anthropology in the recovery and analysis of Branch Davidian compound victims: recovery procedures and characteristics of the victims. *Journal of Forensic Science*, 40(3):335-340.

Wade BS, Joshi SH, Gutman BA, Thompson PM. 2017. Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods. *Pattern Recognition*, 63731-739.

Wescott DJ, and Jantz RL. 2005. Assessing craniofacial secular change in American blacks and whites using geometric morphometry. *Modern morphometrics in physical anthropology*. p 231-245: Springer.

8. DISSEMINATION OF RESEARCH FINDINGS

Workshop and symposia summaries, published abstracts, and reprints are provided as separate .pdf files:

2015-DN-BX-K010_final_report_01_workshops_symposia.pdf

TITLE (WORKSHOP): *New Methods in Skeletal Age Estimation for Diverse Populations (Gr. CONF-772)*

ORGANIZERS: Bridget F.B. Algee-Hewitt and Jieun Kim

LOCATION: Stanford University and Live-Streamed via Webcasting

DATE: August 5-11, 2018.

TITLE (SYMPOSIUM): *Thinking Computationally About Forensics*

ORGANIZERS: Bridget Algee-Hewitt (Chair), Jieun Kim (Co-chair).

LOCATION: The 87th Annual Meeting of the American Association of Physical

Anthropologists, Austin, TX

DATE: April, 2018.

TITLE (WORKSHOP): *A Computational Framework for Skeletal Age-at-Death Estimation Using Laser Scans of the Adult Pubic Symphysis: Theory, Methods, and Software.*

ORGANIZERS: Bridget F.B. Algee-Hewitt (Chair), Jieun Kim (co-chair).

LOCATION: The 67th Annual Meeting of American Academy of Forensic Sciences, New Orleans, LA

DATE: February 2017 (Meeting program: 47).

2015-DN-BX-K010_final_report_02_abstracts.pdf

Kim, J., B. F. B. Algee-Hewitt, D. K. Stoyanova, C. Figueroa-Soto, D. E. Slice. 2019. Testing the applicability of shape-based computational age-at-death estimation methods using pubic symphyseal surface scans of Asian origin. Accepted for the 88th Annual Meeting of the American Association of Physical Anthropology. March 27-30. Cleveland, Ohio.

Figueroa-Soto, Cristina, Jieun Kim, Detelina Stoyanova, Dennis E. Slice, and Bridget F. B. Algee-Hewitt. "Understanding Population Variability in Age-at-Death Estimation for Modern Populations in Mexico and Puerto Rico through the Use of 3D Laser Scans of the Pubic Symphysis." *American Journal of Physical Anthropology* 165 (April 2018): 84–85.

Kim, J., D. Stoyanova, B.F.B Algee-Hewitt, and D. Slice. 2018. Analysis of Inter- and Intra-Observer Error Associated with the Use of 3D Laser Scan Data of the Pubic Symphysis. Paper and poster presented at: (1) the 2018 NIJ Forensic Science Research and Development Symposium; and (2) the 70th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, February, 2018.

Stoyanova, D. K., J. Kim, C. Figueroa-Soto, D. E. Slice, and B. F. B. Algee-Hewitt. 2018. "Age-at-Death Estimation Based on the Female Pubic Symphysis Using Computational Methods and 3D Laser Scans." *American Journal of Physical Anthropology* 165 (April 2018): 266–266.

Figueroa-Soto, C., Stoyonova, D., Kim, J., Slice, D., Algee-Hewitt, B.F. 2017. Testing computational age estimation methods using laser scans of the adult pubic symphysis on modern Hispanic populations. Poster presentation. 2017. American Academy of Forensic Science Annual Meeting. New Orleans, Louisiana.

Kim, Jieun. 2017. "Understanding Population-Specific Age Estimation Using Documented Asian Skeletal Samples." *American Journal of Physical Anthropology* 162 (April 2017): 243–243.

Stoyanova, D. K., B. F. B. Algee-Hewitt, J. Kim, and D. E. Slice. 2017 "Left or Right Pubic Symphysis: Asymmetry Analysis of Age-at-Death Estimation Using 3D Laser Scans and Computational Algorithms." *American Journal of Physical Anthropology* 162 (April 2017): 372–372.

Stoyanova, D., B. F. B. Algee-Hewitt, J. Kim, and D. E. Slice. "A Computational Method for Age-at-Death Estimation Based on the Surface and Outline Analysis of 3D Laser Scans of the Human Pubic Symphysis." 2016. *American Journal of Physical Anthropology* 159 (March 2016): 305–305.

2015-DN-BX-K010_final_report_03_reprint_kim_et_al_2018.pdf

Kim, J., B. F. B. Algee-Hewitt, D. K. Stoyanova, C. Figueroa-Soto, and D. E. Slice. 2018. "Testing Reliability of the Computational Age-At-Death Estimation Methods between Five Observers Using Three-Dimensional Image Data of the Pubic Symphysis,." *Journal of Forensic Sciences* 0, no. 0. <https://doi.org/10.1111/1556-4029.13842>.

2015-DN-BX-K010_final_report_04_reprint_stoyanova_et_al_2015.pdf

Stoyanova, Detelina, Bridget F.B. Algee-Hewitt, and Dennis E. Slice. 2015. "An Enhanced Computational Method for Age-at-Death Estimation Based on the Pubic Symphysis Using 3D Laser Scans and Thin Plate Splines." *American Journal of Physical Anthropology* 158, no. 3 (November 1, 2015): 431–40. <https://doi.org/10.1002/ajpa.22797>.

2015-DN-BX-K010_final_report_05_reprint_stoyanova_et_al_2017.pdf

Stoyanova, Detelina K., Bridget F. B. Algee-Hewitt, Jieun Kim, and Dennis E. Slice. 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." *Journal of Forensic Sciences* 62, no. 6 (November 2017): 1434–44. <https://doi.org/10.1111/1556-4029.13439>.

2015-DN-BX-K010_final_report_06_reprint_stoyanova_et_al_2018.pdf

Stoyanova, D. K., B. F. B. Algee-Hewitt, J. Kim, and D. E. Slice. 2018. "A Study on the Asymmetry of the Human Left and Right Pubic Symphyseal Surfaces Using High-Definition Data Capture and Computational Shape Methods." *Journal of Forensic Sciences* 0, no. 0. <https://doi.org/10.1111/1556-4029.13871>.

2015-DN-BX-K010_final_report_07_preprint_algee-hewitt_et_al_2018.pdf

Algee-Hewitt, Bridget F.B.; Kim, Jieun; and Hughes, Cris E., "Thinking

Computationally about Forensics: Anthropological Perspectives on Advancements in Technologies, Data, and Algorithms" (2018). Human Biology Open Access Pre-Prints. 133.

https://digitalcommons.wayne.edu/humbiol_preprints/133

2015-DN-BX-K010_final_report_08_in_press_slice_2019.pdf

Slice, D. E. IN PRESS. The Development and Use of Computational Tools in Forensic Science. Human Biology.

DISSEMINATION OF RESEARCH FINDINGS

WORKSHOPS/SYMPOSIA

2018 Stanford Workshop

TITLE: *New Methods in Skeletal Age Estimation for Diverse Populations* (Gr. CONF-772)

ORGANIZERS: Bridget F.B. Algee-Hewitt and Jieun Kim

LOCATION: Stanford University and Live-Streamed via Webcasting

DATE: August 5-11, 2018. Public Talks on August 6&7, 2018; Participant Discussions on August 8-10, 2018.

SPONSORS: Wenner-Gren Foundation; Center for Comparative Studies in Race and Ethnicity, Center for Spatial and Textual Analysis, Center for Latin American Studies, Humanities Center, and the Office of the Dean of Research at Stanford University.

PUBLICATIONS: Commitment with Elsevier to produce a volume, entitled, *Re-modeling Forensic Skeletal Age: Estimation, Evidence and Research*, edited by Bridget Algee-Hewitt and Jieun Kim; Publication deadline of January 2020.

Workshop Summary:

The reliable estimation of age-at-death and age-of-the living from the human skeleton is of fundamental importance. It contributes to the study of morphological variation and the senescent process in modern population biology. As such, skeletal development and degeneration has implications that reach into the anatomical and medical fields as we, as skeletal biologists, can grapple with the effects of genetics, environment and individual lifestyle factors on bone density, rates of fracture healing, and atypical skeletal expression. Estimating age is important among the personal identity parameters that are used in medico-legal case identification in forensic anthropology, as we seek to provide law enforcement, medical examiner, and non-governmental agencies with the information that can help to link the unknown individual with the named person. In these circumstances, the anthropologist, working in the service of

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

humanitarian aid and social justice, estimates skeletal age to assist in missing person, asylum seeker, and undocumented death cases. Finally, age-at-death is fundamental to the paleodemographic reconstruction of mortality profiles for skeletal assemblages in bioarcheology, allowing us to bring a better understanding of the life and death of past peoples. Despite the importance of knowing about skeletal “age” and “aging”, achieving accurate, precise, and repeatable estimation continues to be a challenge. Several acknowledged but unresolved theoretical and methodological problems continue to actively constrain the optimal inference of skeletal age for both juveniles and adults.

Our week-long workshop, *New Methods in Skeletal Age Estimation for Diverse Populations*, was supported by the Wenner-Gren Foundation, two co-sponsoring entities, the Center for Spatial and Textual Analysis (CESTA) and the Center for Comparative Studies in Race and Ethnicity (CCSRE), and other contributors at Stanford University, including the Center for Latin American Studies (CLAS), the Humanities Center, and the Dean of Research. This workshop was motivated by the challenges and importance that age estimation holds in not only the traditional field of physical anthropology but also in the new biological anthropology – as we seek out innovative research that supports collaboration and crosses disciplinary, theoretical and methodological barriers. We had the pleasure to invite an esteemed panel of experts in the field of age estimation to participate in the Workshop at Stanford University. Together, we participated in an intensive period of discussion, debate, and even some critical resolution on age estimation practice in the context of forensic casework, modern skeletal research, and the determination of age distributions for diverse populations.

We identified, and we believe successfully realized, three specific aims over the course of this workshop. First, as a broadly experienced community of participants, it was our goal to more fully articulate our current areas of need. Some of which included identifying alternative age-informative traits and new data collection resources, implementing more sophisticated computational techniques, developing models that are more satisfying, statistically, and accessible, practically, obtaining more comprehensive data for diverse populations that better reflect the true range of variation in aging, advancing methods that deliver more realistic mortality profiles for understudied populations, and finally, for the forensic sciences, producing better individual estimates of age and accounting for the evolving evidentiary standards in

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

casework. Second, it was our hope that this workshop would provide a forum to discuss the future of age estimation. The semi-public talks delivered over the first two days represented the innovations in scholarship and the critical re-appraisals of the state-of-the-art of age estimation that are necessary to moving the field forward. Third, the workshop was designed with the aim of motivating the participants, as young thinkers or leaders in the field, and pioneers in their areas of research specialty, to create a set of recommendations on age estimation that defines this future pathway for theory, method, and practice.

We believe, owing to a community of varied thinkers, the workshop group was able to bridge the long acknowledged disconnect between academic and applied work. Articulating our shared – same but different – experiences, we gained fresh insight into how we can modify our computational methods and biological theory to better the anthropologist's aid to the acute needs of changing demographics, humanitarian crises and disaster response. In their keynote talks, Lyle Konigsberg spoke on methods and models for age estimation and Eugenia Cunha delivered practical guidelines for the routine practice in Forensic Anthropology, with an emphasis on some of the challenges particular to practice in Europe. These two talks together embodied the spirit of this workshop as they linked for us both theory and practice. All Monday and Tuesday talks were open to the Stanford/public community as well as invited “virtual” guests, as we live-streamed the talks, questions and discussion periods.

The two-day series of public presentations and keynote talks coupled with the three days of intimate and intensive debate among workshop participants generated a wealth of information for our collaborative volume, entitled *Re-modeling Forensic Skeletal Age: Estimation, Evidence and Research*, edited by Bridget Algee-Hewitt and Jieun Kim. There are limited texts directed at both the researcher and practitioner for age estimation in forensic contexts. Collectively this book is formatted in an innovative manner, presenting highly targeted, hotly debated and critical topics inspired from major knowledge gaps in the field and written for a global perspective by leading, international experts. We present a multidisciplinary approach for age estimation which incorporates groundbreaking research in the fields of data science, systems engineering, forensic anthropology and advanced imaging, which aims to bridge the gap between research, academia and practice. This approach is made accessible by the provision of an end-to-end example from data acquisition, statistical

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

modelling and interpretation using a GUI, as a demonstration for best practice. We provide an introduction, conceptual understanding and taxonomy of statistical frameworks and computational approaches, including the Bayesian paradigm and machine learning techniques for age estimation. We will discuss core concepts in age estimation, key terminology and challenge how we ask questions and generate sound models which can be translated into forensic reports and expert testimony. We provide a step-wise approach and series of recommendations of best practice for data acquisition, considerations in sampling, exploratory data analysis, visualization and sources of error for appropriate and reproducible research design. Specifically, we try to understand the fundamental biological processes of growth and senescence and the impact of different factors for age estimation and discuss the viability of skeletal and dental indicators using micro/macroscopic techniques. This book provides examples, theory and guidance to develop models for age estimation and discuss the impact of population-specific and universal approaches. We provide paradigms for the evidentiary statement and we identify from our collective experience key communication considerations for the presentation of an age estimate for research and the forensic anthropologist. We are confident that this book – emerging directly out of the present workshop – will not only offer a window into exciting, high-level research but also provide a codified set of best practice recommendations made by experts whose knowledge cuts widely across the discipline of biological anthropology, as broadly defined. This workshop product will set the seed for initiating change within the field of skeletal age estimation.

Workshop Participants and Institutional Addresses:

Bridget F.B. Algee-Hewitt, Center for Comparative Studies in Race and Ethnicity, Stanford University, CA.

Guillermo Bravo Morante, Laboratory of Anthropology, University of Granada, Spain.

Hugo Cardoso, Department of Archaeology, Simon Fraser University, Canada.

Louise Corron, Department of Anthropology, University of Nevada – Reno, NV.

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

Eugenia Cunha, Laboratory of Forensic Anthropology, Centre For Functional Ecology, Department of Life Sciences, University of Coimbra, Portugal.

Cristina Figueroa Soto, Office of The Medical Examiner, Waukesha County, WI.

Susan R. Frankenberg, Department of Anthropology, University of Illinois At Urbana-Champaign, IL.

Michael Holton Price, Santa Fe Institute, Santa Fe, NM.

Jieun Kim, Department of Scientific Computing, Florida State University, Tallahassee, FL.

Lyle W. Konigsberg, Department of Anthropology, University of Illinois at Urbana-Champaign, IL.

Nicolene Lottering, Adelaide Medical School, University of Adelaide, Australia.

Marta San-Millan, University School of Health and Sport, University of Girona, Spain.

George R. Milner, Department of Anthropology, Pennsylvania State University, PA.

Stephan Naji, Department of Anthropology, New York University, NY.

David Navega, Department of Life Sciences, University of Coimbra, Portugal.

Kyra Stull, Department of Anthropology, University of Nevada – Reno, NV; University of Pretoria, South Africa.

Chiara Villa, Department of Forensic Medicine, University of Copenhagen, Denmark.

Regular Workshop Talks (with coauthors)

Introduction to Age Estimation. Bridget F.B. Algee-Hewitt

Predicting Age-At-Death From The Shape Of The Human Pubic Symphysis By Bandpass Filtering Of Bending Energy. Guillermo Bravo Morante, Fred Bookstein, Katrin Schaefer, Dennis Slice, Immaculate German Aguilera & Miguel Botella López.

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

Population Variability In Age-At-Death Estimation For Modern Populations In Latin America Through The Use Of 3d Laser Scans Of The Pubic Symphysis. Cristina Figueroa Soto, Jieun Kim, Detelina Stoyanova. Dennis E. Slice & Bridget F.B. Algee-Hewitt.

Acetabular Aging: New Approaches. Marta San Millan, Carme Rissech & Daniel Turbón.

Standardized Cad Approaches To Refine Standards For Age Estimation In Australia Using Thin-Slice Msct: Strengths And Challenges Of Clinical And Post-Mortem Acquisition. Nicolene Lottering.

Going Back To The Basics In Juvenile Age Estimation. Hugo Cardoso.

Exploring The Variation In Growth And Development And Its Impact On Subadult Age Estimation: Variables, Samples, And Models. Kyra Stull, Louise Corron & Michael Price.

Introducing Yada: An Open Source R Package For Demographic Analysis. Michael Holton Price.

Are You A Lumper Or A Splitter? Establishing Age-At-Death Estimation Standards For Two Geographically Distanced Asian Populations. Jieun Kim & Bridget F.B. Algee-Hewitt.

Bayesian Cementochronology...Let's Try It! Stephan Naji

Adult Skeletal Age Estimation: Dismal Results, Ensuing Misconceptions, But A Promising Future. George R Milner, Jesper L Boldsen, Stephen D Ousley, Sara M Getz, Svenja Weise & Peter Tarp.

Testing For Inter-Sample Differences In Senescence. Susan R. Frankenberg & Lyle W. Konigsberg

[Machine] Learning About Age From The Skeleton. Bridget F.B. Algee-Hewitt & Jieun Kim.

Skeletal Age-At-Death Estimation - A Machine Learning Approach. David Navega

Aging The Skull Beyond The Sutures And Living More Than 100 Years: Can We See It On The Skeleton? Eugenia Cunha

New Approaches To Estimate Adult Age-At-Death Using Medical Imaging Techniques. Chiara Villa

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

Keynote Workshop Talks

"It's Not A Scandal. It's Statistics." Lyle W. Konigsberg

Practical Guidelines For The Routine Practice In Forensic Anthropology, With An Emphasis On Some Of The Challenges Particular To Practice In Europe. Eugenia Cunha

2018 AAPA Symposium

Thinking Computationally About Forensics

Bridget Algee-Hewitt (Chair), Jieun Kim (Co-chair). the 87th Annual Meeting of the American Association of Physical Anthropologists, Austin, TX, April, 2018.

ABSTRACT

Computational methods offer several advantages to the study of anthropological data, particularly in their important practical contributions to human identification in the forensic sciences. Through the analysis of large quantities of information, they allow researchers to perform more comprehensive or deeper investigations, effectively overcoming the limitations of cognitive ability and building stronger scientific foundations for applied techniques. By probing data in previously unavailable ways, computational tools also give means to reveal latent data trends, identify and explore novel questions, and establish inferential procedures that deliver more satisfying results. Finally, when computational systems are used to represent expert knowledge, they allow researchers to better capture, distill and interpret complex data, while also improving precision and accuracy, reducing subjectivity, and facilitating the automation of traditional procedures. However, researchers and practitioners alike must contend with evolving issues of software compatibility and data management, bioethical concerns over the new kinds of information now accessible, and the question of best practices for the dissemination of results among peers, in the classroom, for the medico-legal community, and to the public. The purpose of this symposium is to provide a forum to 1) introduce new algorithmic advances and methodological improvement, 2) present work on the application of computational techniques to understudied populations, novel datasets or new information types, and 3) speak to the challenges that the revolution in data technologies may pose for future scientific investigation as well as the broader social effects on issues of policy, privacy and lay interpretation.

This symposium brings together a mix of participants, who engage wide-ranging skeletal, genomic, phenotypic and meta-data analyses. Nevertheless, their contributions are linked by an interest in advancing computational research that has implications for the forensic anthropological sciences, to enrich current

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

procedures and with the potential to change the course of future human identification practice.

2017 AAFS Workshop

A Computational Framework for Skeletal Age-at-Death Estimation Using Laser Scans of the Adult Pubic Symphysis: Theory, Methods, and Software.

Bridget F.B. Algee-Hewitt (Chair), Jieun Kim (co-chair). The 67th Annual Meeting of American Academy of Forensic Sciences, New Orleans, LA, February 2017 (Meeting program: 47).

ABSTRACT

After attending this presentation, attendees will understand state-of-the-art skeletal age estimation, the problems forensic anthropologists face when estimating age by conventional methods, the needs of the medicolegal community, and the potential for advancing the field using new shape-based methods. Attendees will receive instruction in three new methods that apply numerical shape algorithms to laser scans of the skeletal age indicator. Attendees will also receive hands-on training in using the equipment, software, casts, and data. This presentation will impact the forensic science community by delivering instruction on implementing three new fully computational methods for age-at-death estimation from skeletal laser scans that produce estimates that closely approximate true age, with minimal risk of subjectivity or low-method/observer-induced error. The estimation of age-at-death in forensic anthropology represents an essential component of the biological profile, providing information on the individual that is key to medicolegal case identification. Skeletal indicators of age are widely used to estimate age-at-death from adult remains in this casework context. Of the pelvic, thoracic, and cranial features for which age-related change is known, the pubic symphysis remains the preferred, most frequently studied indicator. Common practice requires the macroscopic comparison of the bone surface morphology to a set of population-specific criteria that represent a series of pre-defined scores or phases. The case-specific age-at-death is then estimated from an age range previously associated with the assigned score or phase. While the simplicity of this approach is attractive, the limitations of this kind of visual analysis are well-documented across the field, not only for age-at-death estimation and but also for other parameters of interest to the biological profile. In general, this methodology is known to introduce a large degree of subjectivity and intra/inter observer-related

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

error. For age estimation especially, these problems have been shown to variably impact the reliability and repeatability of results; therefore posing, significant challenges to meeting current medico-legal standards of evidence and successful forensic case identification. In response to these concerns, an alternative, fully computational approach to the macromorphoscopic assessment methods traditionally applied to skeletal indicators of age and, specifically, to the pubic symphysis has been proposed. It is argued that accurate, precise, and objective age estimates can be obtained by sourcing three-dimensional coordinate data from laser scans of the pubic symphysis, subjecting these data to shape-analysis algorithms, and combining the resulting shape measures in multivariate regression models. In recent publications the value of this novel approach for contemporary skeletal analysis have been demonstrated. Using Bass Collection samples, forensic cases, and the Suchey-Brooks and McKern & Stewart casts, these methods produce estimates that differ from the exact age-at-death by $\approx 11.72 \pm 0.97$ years. To standardize implementation, a protocol for data collection and extraction has been formalized. The software, *forAge*, was developed to facilitate accurate and efficient method application among forensic practitioners, whose levels of familiarity with laser scanning technology, statistical computing, and the morphological characteristics of the pubic symphysis may vary. This workshop will introduce forensic practitioners to the theory that underlies the methods and provide laboratory instruction on its implementation. It will also be broadly applicable to other scan and shape-related research. To contextualize the work, a review of the current state of the art of age-estimation, the demands that working within the medico-legal context places on forensic case analysis, and the advantages that these methods offer for estimation, evidence, ease of use, and data preservation or sharing. The anatomical properties of pubic symphyseal morphology that make this indicator well-suited to shape-based inference as well as the mathematical theory that supports the calculation of our shape measures will be explained. The appropriate use of these measures 18 *Presenting Author and clarification of how an estimate of age is generated will be discussed. To provide practical instruction, protocols for laser scan collection using the NextEngine scanner, scan editing and manipulation using ScanStudio or Meshlab, extraction of shape information as three-dimensional coordinates, file storage of these data, and for standardization and processing of the coordinates prior to analysis will be demonstrated. The detailed implementation of the following will be discussed: (1)

FINAL REPORT: AWARD NO: 2015-DN-BX-K010

the SAH-Score method that captures the variance on the symphyseal face to capture the gradual flattening of the surface associated with aging;¹ (2) the thin plate splines method that determines the bending energy required for transforming a perfectly flat, infinitely thin plate to match the surface of a pubic symphysis scan;² and, (3) the ventral curvature method that quantifies the progressive formation of a rim around the entire symphyseal surface and its later erosion.³ A tutorial on the use of the *forAge* software for these analyses and to produce age-estimates via multivariate regression will be provided. With France Casting, the age-determination casts for calibration and validation will be discussed. Finally, recommendations for data collection in the field and laboratory will be offered. Workshop attendees will have the opportunity to train on the scanning equipment and software and produce age estimates directly from specimens and coordinate data in various stages of processing.

Reference(s):

1. Slice D, Algee-Hewitt, B. 2015. Modeling Bone Surface Morphology: A Fully-Quantitative Method for Adult Age-At-Death Estimation Using the Pubic Symphysis. *J Forensic Sci* 2015: 60(4): 835-843.
2. Stoyanova D, Algee-Hewitt B, Slice D. An Enhanced Computational Method for Age-At-Death Estimation Based on the Pubic Symphysis Using 3D Laser Scans and Thin Plate Splines. *Am J Phys Anthropol* 2015: 158: 431-440.
3. Stoyanova D, Algee-Hewitt B, Kim J, Slice D. 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* 2016: in review (submitted April. 26, 2016).

ABSTRACTS: 2019 (accepted)

Testing the applicability of shape-based computational age-at-death estimation methods using pubic symphyseal surface scans of Asian origin

JIEUN KIM¹, BRIDGET FB. ALGEE-HEWITT^{1,2,4}, DETELINA K. STOYANOVA^{1,3}, CRISTINA FIGUEROA-SOTO⁴ and DENNIS E. SLICE^{1,5}

¹Scientific Computing, Florida State University, ²Center for Comparative Studies in Race and Ethnicity, Stanford University, ³Mathematics and Statistics, University of North Carolina at Charlotte, ⁴Anthropology, University of Tennessee, ⁵Anthropology, University of Vienna

Recent computational age-at-death estimation methods developed using 3D laser scans of the pubic symphysis have been shown to provide robust estimates of age for documented White individuals. While validation testing has demonstrated reduced within-/between-observer error, improved objectivity, invariance to asymmetry, and equal applicability to female and males pubic symphyses, no study has explored the issue of population diversity. Concerns over broad applicability arise from the facts these methods were developed using a reference sample composed of modern American White males and that the same sample is used to produce the final age estimate in the associated software, forAge, which implements the shape algorithms and multiple regression analyses. There is, therefore, a need to determine their utility for peoples from different geographic regions. The present study seeks to assess the applicability of these computational methods for age estimation of Asians, focusing on a mixed-sex sample from Asia. Three shape-based measures, capturing the gradual flattening of the face and changes in the ventral margin of the symphysis, are used to build a series of regression models and the final age estimates are assessed for error and bias. Preliminary results from 69 scans suggest statistically significant relationships exist between ages-at-death and the three shape measures ($p < 0.05$). R-squared values indicate that 30-50% of the shape variation can be explained by age. RMSE values of 10-11 years are lower than those originally reported. These preliminary results suggest the utility of these methods for Asia and support further investigation for the rest of the region.

This project is supported by a National Institute of Justice grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

ABSTRACTS: 2018

Understanding population variability in age-at-death estimation for modern populations in Mexico and Puerto Rico through the use of 3D laser scans of the pubic symphysis

CRISTINA FIGUEROA-SOTO¹, JIEUN KIM², DETELINA STOYANOVA², DENNIS E. SLICE² and BRIDGET FB. ALGEE-HEWITT³

¹Anthropology, The University of Tennessee, ²Scientific Computing, Florida State University, ³Biology, Stanford

One of the main goals of the forensic anthropologist is to aid in the positive identification of unknown human remains by developing a biological profile. Age-at-death estimation is a crucial parameter, but it remains one of the most challenging. Reliable estimation is made difficult by the fact that many traditional methods are dependent upon a set of population-specific criteria that have been mainly developed using individuals of European and African descent. The absence of information on the potential differences in the aging patterns of underrepresented, especially Latino, populations across the U.S. may, therefore, hinder our efforts to produce useful age-at-death estimates. In response to this concern, this study obtained data from individuals of Mexican and Puerto Rican origin to test and update a newly published computational framework for age-at-death estimation from the pubic symphysis for contemporary Hispanic casework. Data for this study consist of laser scans of the pubic symphysis from skeletal collections with known age-at-death at the Universidad Nacional Autónoma de Mexico and the Institute of Forensic Science in Puerto Rico. Each scan was subjected to the Slice and Algee-Hewitt(SAH), Thin Plate Spline/Bending Energy (BE), and Ventral Curvature (VC) methods. Preliminary analyses, using paired t-tests, find no significant differences between known and inferred age-at-death ($0.03 \leq p \leq 0.71$) after Bonferonni correction. Results do suggest a tendency for the VC and BE methods to underestimate age for individuals over 65 years. This study has confirmed that reliable age-at-death estimates can be obtained for Hispanic groups using these computational methods.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 165, 84-85

Analysis of Inter- and Intra-Observer Error Associated with the Use of 3D Laser Scan Data of the Pubic Symphysis.

Jieun Kim, Detelina Stoyanova, Bridget Algee-Hewitt, and Dennis Slice.

ABSTRACT

After attending this presentation, attendees will understand the intraobserver and interobserver error related to collecting and editing 3D laser scans of skeletal material, as well as the repeatability of three new age estimation methods that use bone shape data extracted from the laser scans. This presentation will impact the forensic science community by providing best practice guidelines for using laser scanners, scanned images, and coordinating data in forensic casework that will contribute to the standardization of 3D image processing procedures between different forensic practitioners and labs. It will also provide validation for the age estimation methods discussed here. In age-at-death estimation based on visual assessment, objective evaluation and correct diagnosis of age-related skeletal traits are crucial, as these factors determine whether the aging methods can achieve their full potential — producing the most accurate and reliable age estimates. Nevertheless, the traditional, phase-based age estimation methods have been reported to yield inconsistent age estimates both within and between observers.^{1,2} The reasons for these discrepancies lie in the fact that accurate macromorphoscopic analysis heavily depends on the correct interpretation of qualitative trait descriptions, conformity of the bone, and experience of the observer. Recently, Slice and Algee-Hewitt and Stoyanova et al. have introduced three novel, fully computational aging methods using 3D laser scans of the pubic symphysis that minimize subjectivity in age estimation by reducing the effects of observer experience in the age-indicator/trait assessment and methodological bias; however, the reproducibility of these methods has not been fully explored or quantified.³⁻⁵ This is of concern because there is potential for introducing error in the first two steps of data processing — when the scans are taken and edited at different times by different observers. In response to this concern, the current study evaluates the repeatability of these novel methods by assessing intra-scan variation, within, and between, observer differences in scan editing and its impact on age estimation. The test data used in this study represent replicate scans of the Suchey-Brooks' (SB) male casts, taken using a 3D desktop laser scanner. The upper and lower stages of each of the six phases were scanned three times by a single observer (n=36). Four different observers with various experience levels and training backgrounds independently edited the triplicate of the SB scans using the scanner's accompanying software, such that the symphyseal face is extracted from the

surrounding bones. From these isolated faces, x, y, and z coordinates were retrieved and analyzed via the Subarachnoid Hemorrhage (SAH) Score method, the Thin Plate Splines/Bending Energy (TPS/BE) method, and the Ventral Curvature (VC) method to compute shape measures.³⁻⁵ These measures were subjected to single-variable and multivariate regression models to obtain age estimates for each replicate scan per observer. Finally, using the shape measures and final age estimates, a series of the Intraclass Correlation Coefficient (ICC) were calculated to evaluate within- and between-observer reliability in scan editing. Additionally, extra editing conditions were tested to simulate the situation in which the practitioner misidentifies age-related traits due to unfamiliarity with the scan editing protocol. A set of the SB casts was edited with different widths of the margin (2mm vs. 4mm vs. 1cm) left around the symphyseal face and with/without the pubic tubercle, which may impact the VC values as it protrudes ventrally. Possible effects of these conditions on age estimates were evaluated using the paired t-test. This study produced high ICC values (0.75-1.0), demonstrating that the raw scans were edited consistently within and between observers and that the derived shape measures and age estimates were in excellent agreement among observers. Moreover, despite the simulated improper editing of the scans with various margin widths remaining, the methods were robust enough to self-correct and produce consistent and accurate age estimates ($p > 0.05$), with the exception of the faces with 1cm margin. Interestingly, the inclusion of the pubic tubercle for the shape analysis did not necessarily yield inaccurate age estimates for the VC method, while it produced statistically significant mean differences between the documented chronological age and age estimates of the SAH score method, TPS/BE method, and the two multivariate regression models ($p < 0.01$). These results demonstrate high repeatability of the computational methods regardless of the observer's level of experience or training background and support using a 3D laser scanner and scanned images to aid in resolving the issue of subjectivity.

Reference(s):

1. Kimmerle E.H., Prince D.A., and Berg G.E. Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth. *J Forensic Sci.* 2008;53(3): 594-600.
2. Shirley N.R. and Ramirez Montes P.A. Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component-Based Methods. *J Forensic Sci.* 2015;60(1): 107-111. doi:10.1111/1556-4029.12617.

3. Slice D.E., Algee-Hewitt B.F. Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. *J Forensic Sci.* 2015;60(4):835-43.

4. Stoyanova D., Algee-Hewitt B.F., Slice D.E. An Enhanced Computational Method for Age-at-Death Estimation Based on the Pubic Symphysis Using 3D Laser Scans and Thin Plate Splines. *Am J Phys Anthropol.* 2015;158(3):431-40.

5. Stoyanova D., Algee-Hewitt B.F., Kim J., Slice D.E. A Fully Computational Framework for Age-at-Death Estimation from the Adult Skeleton: Surface and Outline Analysis of Three-Dimensional Laser Scans of the Pubic Symphysis. *J Forensic Sci.* 2017. doi:10.1111/1556-4029.13439

Paper and poster presented at: (1) the 2018 NIJ Forensic Science Research and Development Symposium; and (2) the 70th Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, February, 2018.

Age-at-death estimation based on the female pubic symphysis using computational methods and 3D laser scans

DETELINA K. STOYANOVA^{1,2}, JIEUN KIM¹, CRISTINA FIGUEROA-SOTO³, DENNIS E. SLICE^{1,4} and BRIDGET FB. ALGEE-HEWITT^{1,5}

¹Scientific Computing, Florida State University, ²Department of Mathematics and Statistics, University of North Carolina at Charlotte, ³Waukesha County Medical Examiner's Office, ⁴Department of Anthropology, University of Vienna, ⁵Department of Biology, Stanford University

The most popular and widely used skeletal indicator for age-at-death estimation is the pubic symphysis. Recently, novel computational, shape-based, aging methods have been proposed as alternatives to traditional visual-scoring methods to address recognized limitations, especially high inter-/intra-observer error. However, these new methods were developed on 3D laser scans of males, and their applicability to female skeletal remains has not been substantiated. In this study, we show that these computational techniques can successfully be applied to females despite extrinsic factors (e.g., childbirth and osteoporosis) that may alter the morphology of the pubic symphysis and make accurate age-at-death assessment of females challenging. This study uses 3D laser scans from 55 documented female skeletons whose ages range from 15 to 99 years. The results are based on two surface scores and one outline measure that are shown to be associated with age in males. As our sample is skewed towards older individuals, we generate different data subsamples, preserving the ratio between individuals who are younger and older than 40 years. The results of regression models produced for the different subsamples show that the shape measures are associated with recorded age-at-death. R-squared values indicate that 30% to 55% of the shape variation is explained by age and p-values are significant ($\alpha=0.05$). An additional individual asymmetry analysis based on 34 females shows that both sides can be used interchangeably for age-at-death estimation using these new methods. Furthermore, testing finds that individual asymmetry is not significantly associated with advanced age, number of children, weight or stature.

This project was supported by a National Institute of Justice grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 165, 266-266

ABSTRACTS: 2017

Testing computational age estimation methods using laser scans of the adult pubic symphysis on modern Hispanic populations.

Figueroa-Soto, C., Stoyanova, D. Kim, J. Slice, D. Algee-Hewitt, B.F.

After this presentation, attendees will have a better understanding of the applicability and effectiveness of three newly published computational methods that focus on 3D laser scans of the pubic symphysis (Slice and Algee-Hewitt 2015; Stoyanova et al. 2015; Stoyanova et al. 2016) for the estimation of age at death on populations of Mexican and Puerto Rican descent.

This presentation will impact the forensic science community by demonstrating how age at death can be estimated in a more accurate, precise and objective way by utilizing fully computational methods and 3D scans of the pubic symphysis on populations of Mexican and Puerto Rican descent. It will also provide recommendations and best practice applications of these standards to forensic casework of a possible Hispanic background.

Age-at-death estimation techniques have received considerable attention within the anthropological community, especially among forensic anthropologists, as it can narrow the list of potential missing persons in a forensic investigation. Even though the estimation of age is a crucial parameter of a biological profile, it is also one of the most challenging to attain as it greatly depends on the practitioner's ability to associate age changes against a set of population specific criteria, usually represented by a series of pre-defined age phases. Along with this issue is the lack of population specific standards for underrepresented populations in the U.S., as most of the age-at-death techniques are based on individuals of African and European ancestries from late 19th and mid-20th century anatomical collections. In order to address the lack of population specific methods in age at death estimation, this study sourced data from individuals of Mexican and Puerto Rican decent with the goal to test the applicability of newly published standards for age at death estimation created by Slice and Algee-Hewitt¹, and Stoyanova et al^{2, 3}.

Slice and Algee-Hewitt, and Stoyanova et al. have recently developed a more objective, fully computational, and statistically robust technique than conventional bone-to-phase matching standards. These new techniques utilize coordinates obtained from 3D scans of the pubic symphysis that are subjected to numerical shape algorithms and regression analysis producing objective age-at-death estimates. The ultimate goal of this study is to expand these new standards so that they can be utilized in populations of Mexican and Puerto Rican ancestry and, in turn, to produce more accurate and precise estimates of age at death with reduced

error and subjectivity than currently possible using the traditional macroscopic assessment methods.

Data for this study consist of laser scans for both sides of the pubic symphysis from skeletal collections with known age-at-death housed at the Universidad Nacional Autónoma de México, Universidad Autónoma de Yucatán, Institute of Forensic Science in Puerto Rico, and Pima County Office of the Medical Examiner's in Arizona. For each pubic symphysis, 3D scans were created using the NextEngine 3D Desktop Scanner, 2020i. The resulting scans were aligned and fused in order to generate a multidimensional model of the pubic symphysis. The 3D coordinates of each scan were extracted and subjected to the *SAH-Score* method¹, the thin plate splines/bending energy-based method², and the ventral curvature method³. Multivariate regression models were used to combine the resulting measures and obtain the final age estimates for each individual.

Preliminary results demonstrate no significant differences between the known age-at-death and the inferred age on all three methods, when both sides, left and right, were tested separately and together. Furthermore, there is a tendency, though not significant, to overestimate known age-at-death for all methods, with mean differences as low as -9.3 and as high as -1. Overall, we have demonstrated that objective and reliable age-at-death estimation can be obtained on populations of Mexican and Puerto Rican descent by the use of computational methods and 3D laser scans of the pubic symphysis.

References:

¹ Slice DE, Algee-Hewitt BF. Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. *Journal of forensic sciences*. 2015;60(4):835-43.

² Stoyanova D, Algee-Hewitt BF, Slice DE. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *American journal of physical anthropology*. 2015;158(3):431-40.

³ Stoyanova D, Algee-Hewitt BF, Kim J, Slice DE. A Fully Computational Framework for Age-at-Death Estimation from the Adult Skeleton: Surface and Outline Analysis of Three-Dimensional Laser Scans of the Pubic Symphysis. *Journal of forensic sciences*. 2016.

Poster presentation. 2017. American Academy of Forensic Science Annual Meeting. New Orleans, Louisiana.

Understanding population-specific age estimation using documented Asian skeletal samples. Invited Poster Symposium: Broadening Forensic Anthropology: Bringing East and Southeast Asia to the Forefront.

Jieun Kim

ABSTRACT

Understanding aging patterns and rates of the adult skeleton is one of the most challenging tasks to skeletal biologists because senescence is easily influenced by both intrinsic (i.e., genetics) and extrinsic (i.e., nutrition, pathology, activities, etc.) factors. Among the myriad of different variables, ancestral affiliation has been suggested to play a highly impactful role in influencing age-related trait expressions, such that applying one method developed on a particular population is not relevant to another population. While there has been a strong emphasis on developing and using population-specific methods, what level of population-specificity one age estimation method should offer is relatively underexplored. The key to addressing this question lies in a holistic understanding of population history as well as systematic studies of subgroups within a continent. This study investigates whether an aging method should be region-specific or continental-specific using 20th century Japanese (n= 183) and modern Thai (n=236) individuals from four documented skeletal collections. Four age estimation methods were applied to the Asian samples, including Transition Analysis (Boldsen et al. 2002), and three conventional methods (Suchey and Katz 1998, Lovejoy et al. 1985, Meindl and Lovejoy 1985). Final age estimates were obtained using multivariate ordered probit regression under Bayesian inference. The results show that an age estimation model derived from a pooled-Asian sample performs superior to Japanese-/Thai-specific models. In addition, error and bias in age estimates induced by biased reference samples are greater than expected, and such error can be falsely interpreted due to between-population variation in skeletal aging.

The 86th Annual Meeting of the American Association of Physical Anthropologists, New Orleans, LA, April 2017 (Meeting Program:109)

Left or Right Pubic Symphysis: Asymmetry Analysis of Age-at-Death Estimation Using 3D Laser Scans and Computational Algorithms.

Detelina Stoyanova¹, Bridget Algee-Hewitt^{1,2,3}, Jieun Kim^{1,3} and Dennis Slice^{1,4}

¹Department of Scientific Computing, Florida State University, ²Department of Biology Stanford University, ³ Department of Anthropology, Florida State University, ⁴Department of Anthropology, University of Vienna

Age-at-death estimation is crucial for building individual forensic profiles and studying mortality in past populations. For decades, anthropologists have relied on imprecise age estimation techniques based on the visual inspection of the pubic symphysis. Recently several computational methods using 3D laser symphyseal scans have been proposed as accurate, reliable and objective alternatives to current practices. The methods include two surface analysis algorithms, one ventral outline measure, and two multivariate-regression models combining each surface measure with the outline score. The five proposed models are calibrated on 3D scans from white males where the left or right pubic symphysis was randomly selected. A question remains whether the asymmetry of the two surfaces affects the age estimates. For this study both the left and right pubic symphyses from 25 white males are scanned. Both sides are used to estimate the age-at-death for each individual using the five computational models. Additional tests are performed by selecting 25 individuals for whom the left or right symphyseal scan is included in the data for the models. For those males the age-at-death is estimated using the opposite side. The results of paired t-tests for mean differences, Wilcoxon rank sum tests for median differences and Kolmogorov-Smirnov tests for distributional differences show that there is no significant difference (p-values > 0.26) between the age estimates of the two sides for the 50 males. The Spearman and Pearson correlations are robustly positive, between 0.47 and 0.72 (p-value < 0.05), suggesting a monotonic relationship with high degree of linear dependence.

This project was supported by a National Institute of Justice grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 162, 372-372

ABSTRACTS: 2016

A Computational Method for Age-at-Death Estimation Based on the Surface and Outline Analysis of 3D Laser Scans of the Human Pubic Symphysis

DETELINA STOYANOVA¹, BRIDGET F. B. ALGEE-HEWITT^{2,3}, JIEUN KIM⁴ and DENNIS E. SLICE^{1,5}

¹Scientific Computing, Florida State University, ²Anthropology, Florida State University, ³Biology, Stanford University, ⁴Anthropology, University of Tennessee, ⁵Anthropology, University of Vienna.

Accurate age-at-death estimation is crucial for building individual forensic profiles and studying mortality in past populations. The pubic symphysis is the most widely used skeletal age indicator. In standard practice, symphyseal shape is visually compared to phases, whose morphological characteristics are associated with age intervals. This kind of method introduces some level of subjectivity and observer-related error. Recently two novel, objective techniques have been proposed that use a 3D scan representation of the shape of the pubic symphysis and apply computational algorithms to capture the age-related features of the surface. Both methods use laser scans from ≈ 50 modern American male skeletons with known ages-at-death. The first technique, the SAH-Score, measures the scan's surface variance, while the second uses the bending energy produced by the thin-plate spline algorithm to quantify surface structure. These methods are able to capture the transition of the symphyseal face from being covered by well-developed ridges and furrows, for younger individuals, to flattening with increased age. Both methods analyze the same feature of the pubic symphysis – the face. We present an alternative algorithm that measures the curvature of the ventral margin of the symphyseal scans. When applied to the original data used by the two surface analysis techniques, this new measure shows comparable results: RMSE=18.3 years and R-squared=0.19. However, when the face and margin measures are combined in a multivariate regression model, there is a RMSE improvement of about 2 years and an improvement in R-squared of over 10%.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 159, 305-305

TECHNICAL NOTE**ANTHROPOLOGY**

Jieun Kim ^{1,†} Ph.D.; Bridget F.B. Algee-Hewitt,^{1,2,3,‡} Ph.D.; Detelina K. Stoyanova,^{1,4} Ph.D.; Cristina Figueroa-Soto,^{3,5} M.A.; and Dennis E. Slice,^{1,6} Ph.D.

Testing Reliability of the Computational Age-At-Death Estimation Methods between Five Observers Using Three-Dimensional Image Data of the Pubic Symphysis*[†]

ABSTRACT: In an effort to standardize data collection and analysis in age estimation, a series of computational methods utilizing high-dimensional image data of the age indicator have recently been proposed as an alternative to subjective visual, trait-to-phase matching techniques. To systematically quantify the reproducibility of such methods, we investigate the intrascan variability and within- and between-observer reliability in initial scan data capturing and editing using 3D laser scans of the Suchey–Brooks pubic symphysis casts and five shape-based computational methods. Our results show that (i) five observers with various training background and experience levels edited the scans consistently for all three trials and the derived shape measures and age estimates were in excellent agreement among observers, and (ii) the computational methods are robust to a measured degree of scan trimming error. This study supports the application of computational methods to 3D laser scanned images for reliable age-at-death estimation, with reduced subjectivity.

KEYWORDS: forensic science, forensic anthropology, age-at-death estimation, observer error, 3D laser scans, pubic symphysis, biological profile, forensic casework, osteological standards, morphometrics, multiple regression

Age-at-death for the adult skeleton remains one of the most challenging biological profile parameters requiring osteological estimation. This difficulty arises, first, from the fact that aging, or the postmaturation degeneration, of the adult skeleton is a complex process that reflects dynamic interplays between genetic and nongenetic (i.e., hormonal, biomechanical stress, nutritional, disease, body size) factors that are unique to each individual (1–3). These factors determine how age-related skeletal characteristics manifest in the individual and consequently cause a wide range of variation in skeletal trait expression between individuals and

between age-informative indicators within the individual. Accordingly, they contribute to the apparent delay and/or acceleration of skeletal aging, which results in an imperfect correlation between chronological age and biological age (4,5). Each skeleton can be said, therefore, to have its own degree of error (6,7). Reliable age estimation is further challenged by the methodological error associated with the traditional morphological assessment systems. Indeed, deficiencies in the most popular and preferred macromorphoscopic (Ousley and Hefner (2005) after Wilczak and Christopher [8]) (morphological variation) methods have been already well articulated and include issues related to the bone-to-phase matching scoring systems (5,6,9), the weak statistical foundations for data analysis and age inference (10–12), the skewed reference samples used in method development (13,14), the reliance upon subjective, qualitative trait descriptions (15), and the prominent role that the experience of the observer plays in the successful application of most age estimation methods (15,16).

Over the past three decades, alternative approaches to age estimation have been proposed, with the goal to overcome the limitations associated with traditional methods. Out of this work has emerged a consensus on what properties make for good age estimation methods: they should use revised, more quantitative and objective skeletal trait evaluation systems and nontraditional traits distributed throughout the skeleton (7,17–20), more robust statistical frameworks that can take into account the target population's current mortality profiles (6,21–27), and expanded reference skeletal samples incorporating bio-geographically diverse populations (28). In response to these concerns and owing to the recent technological advancements in scientific computing and data capture

¹Department of Scientific Computing, Florida State University, Tallahassee, FL 32306.

²Department of Biology, Stanford University, Stanford, CA 94305.

³Department of Anthropology, University of Tennessee, Knoxville, TN 37996.

⁴Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223.

⁵Waukesha County Medical Examiner's Office, Waukesha, WI 53188.

⁶Department of Anthropology, University of Vienna, Vienna 1090, Austria.

Corresponding author: Jieun Kim, Ph.D. E-mail: jkim17@fsu.edu

*Supported by a National Institute of Justice grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

[†]Authors contributed equally.

[†]Presented at the 2018 National Institute of Justice (NIJ) Forensic Science Research and Development Symposium, February 20, 2018, in Seattle, WA; and at the 70th Annual Scientific Meeting of the American Academy of Forensic Sciences, February 19–24, 2018, in Seattle, WA.

Received 13 Feb. 2018; and in revised form 27 April 2018, 29 May 2018; accepted 5 June 2018.

techniques, there has emerged a body of innovative research that utilizes computerized, virtual age indicators and imaging data (e.g., CT/MRI scans) to expand the scope of the analysis of age-progressive morphologies, enabling for the first time fully quantitative approaches to the analysis of skeletal data and the estimation of chronological age (29–38). In addition to increasing accuracy and precision in age estimation, these new approaches seek to improve the degree of repeatability and the level of standardization in both data collection and analysis, with an explicit interest in responding to the evidentiary needs of the medico-legal community in the context of forensic anthropological casework.

Recent work by Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41) represents successful examples of studies that embrace this methodological shift in adult age estimation. In these papers, the authors have proposed an alternative “fully computational” way of evaluating age-related skeletal traits by applying a series of shape-based algorithms to three-dimensional laser scans of the pubic symphysis that yield continuous data measures, which can be used in predictive models. Specifically, the *Slice-Algee-Hewitt (SAH)* score and Bending Energy (BE) capture the changes in the complexity (or gradual flattening) of the pubic symphyseal surface. The *SAH* score measures the variance on the surface of the symphyseal face from the third eigen value of a principal component analysis: the greater the variance (or *SAH* score), the younger the individual. The BE value, obtained through the thin plate splines (TPS) algorithms, represents the minimal energy required to bend an infinitely thin flat plate to match the surface of the pubic symphysis. As such, young individuals with more complex surface topography yield greater BE values. Lastly, the ventral curvature (VC), measured by finding a best-fitting circle through the semi-landmarks of the ventral outline, quantifies the changes in the ventral margin of the symphyseal face throughout the adulthood. A large VC value indicates that the pubic symphysis is from an older individual. The steps for implementing these methods proceed such that (i) the pubic symphysis is scanned; (ii) the observer edits the scan to isolate the symphyseal face from the rest of the pubic bones; (iii) three-dimensional coordinates are extracted from the edited scans; and (iv) the coordinates are subjected to numerical shape algorithms and regression analysis to attain final age estimates. The analysis, in step 4, can be executed using the free Java-based software, *forAge* (<http://morphlab.sc.fsu.edu/software/forAge/index.html>), which was designed to calculate easily for the user both the shape-based scores and the final age estimates based on simple and multiple regression models.

The method development studies by Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41) have demonstrated how the three sets of shape-based measures yield accurate and reliable age estimates that are comparable (sometimes even better performing) to those of conventional aging methods, namely the Suchey–Brooks system. Moreover, their ongoing validation studies, based on a large-scale, mixed sex, and multi-ethnic sample, continue to show promising, albeit preliminary, results: for example, the shape methods are invariant to issues of asymmetry in the pubic symphysis and can be successfully applied to males and females of various geographic origins and ancestral histories (42–44). From the success of these studies, the authors contend that this semi-automated procedure improves age estimation practice by reducing the effects of observer experience and/or methodological bias. However, there remain concerns over method reproducibility as there is the potential for introducing error in the first two steps of data processing, when the scans are taken and edited at different times by possibly different observers.

While Stoyanova et al. (40) have shown the increased repeatability of the fully computational methods through a single (intra) observer error test, their methods’ performance and reliability have not been fully quantified or substantiated at the level of multiple observers with different training backgrounds and/or different experience levels. Moreover, the potential for introducing error during scan editing arguably increases if the observer is (i) unfamiliar with the editing protocols of raw scans, (ii) inexperienced with skeletal biology and/or age estimation, and/or (iii) provided with scans of poor resolution or scans with no surface color/texture of the bone captured. Specific to the application of these computational methods to pubic symphysis scans, there are two regions of the bone where editing mistakes and observer variation are more likely to occur: the margins of the pubic symphysis and the pubic tubercle protruding on the ventral aspect of the bone. Trimming around these two areas could be challenging, especially when (i) the pubic symphysis exhibits ridges and billows covering the symphyseal face in part/entirely, (ii) there are bony outgrowths or instances of lipping close to/at the margin, and (iii) the pubic tubercle extends to and is integrated into the superior aspect of the symphyseal face. All these conditions may obscure proper identification of the region of interest (the features required to properly execute the methods) relative to the remaining bone area. For example, the inclusion or exclusion of the pubic tubercle can potentially impact the outcome of the VC algorithm by making it fit circles with different curvature values.

Because discrepancies can arise from such various sources, we argue here that measuring the degree of error between observers at the preprocessing stage—during data capture via scanning and when editing the resultant images—is a critical issue. It demands attention not only for the optimal implementation of these computational age estimation methods, but also for any methods that utilize high-dimensional images as a source of numerical data. The goal of this study was, therefore, to test the following research hypotheses:

- Hypothesis 1. The same observer will edit a set of raw scans inconsistently when editing is done repeatedly over time.
- Hypothesis 2. Observers with various training background and/or level of experience will edit the same sets of raw scans inconsistently.
- Hypothesis 3. Edited scans with margins of different widths left around the pubic symphyseal face will yield different shape measures and age estimates.
- Hypothesis 4. Edited scans with the ventrally protruding pubic tubercle will yield inaccurate age estimates when applying the VC method.

This extension work, which builds upon the method development and validation studies of Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41), is necessary to determine the effect of inter-/intraobserver and experience-related error on performance—specifically, the potential to produce the most accurate, reliable, and precise age estimates when implementing the fully computational methods as proposed in the original papers.

Materials and Methods

Study Sample

The test data used here represent replicate scans of the 12 Suchey–Brooks’ (SB) age estimation male casts, taken by the same make of portable, desktop laser scanner, the NextEngine

3D Scanner HD, as in the original papers. Specifically, it is identical in all specifications to the machine used in Slice and Algee-Hewitt (39) but represents an earlier generation model to that used for the newer scans in Stoyanova et al. (40,41). For this study, the upper and lower stages of each of the six phases were scanned three times by a single observer ($n = 36$). The SB casts were chosen exclusively for the analysis given their accessibility to most researchers, allowing for easy replicability of our analysis. We acknowledge that this gain in reproducibility does come at the expense of testing error for many samples that may cover a wider range of morphological variation.

Each of the SB male casts was placed on the small platter of the AutoDrive scanning stand approximately 6.5 inches away from the scanner. The raw scans were produced within 5 min, using the scanner's recommended high definition (HD) settings: neutral-colored object, 16 divisions, 40 K points per inch² with a triangle size of 0.0050, and a close-up (macro) single scanning mode. The casts' labels were covered and, prior to each scanning, the casts were shuffled to produce a new order. Scanning sessions were broken up by irregular periods to introduce variation in the time between scans. "Blinded" casts were chosen from the "new" order to randomize the scanning at each session. Scans were *post hoc* associated with their true order and labels.

Given the fact that the scan editing protocols and computational methods will most likely be used by researchers who have, at least, minimal experience, knowledge, or training in age estimation and/or human osteology/anatomy, four different observers with common interests and/or specialty in forensic sciences, and yet with various experience levels and training background, were randomly chosen among the pool of potential participants representing method developers, method experts, and method users. A fifth observer was selected to represent the extreme of a novice category of users, as discussed later.

Three of the four main observers independently edited the triplicate sets of the SB scans throughout a 4-week period; Observer 1 took additional time between scan edits to simulate, as best possible, *tabula rasa* conditions. Of the four, three researchers are biological anthropologists and represent a senior researcher (Observer 1), a newly-graduated Ph.D. (Observer 3), and an advanced graduate student/forensic investigator (Observer 4). All three had received training in skeletal biology and have practiced the computational methods prior to this study to untested target samples in North America and South America. Particularly, Observer 1 is the developer of the scan editing protocol for the computational methods and an experienced osteologist. Observer 2 is a computational scientist and the developer of the computational methods and accompanying software, *forAge*, with some knowledge in osteology and pubic symphyseal scan editing experience, largely acquired for the larger method development project. At the time of the study, Observer 1 had almost a decade of scan editing experience and Observer 2 had 5 years of experience with scan editing, while Observers 3 and 4 had 2 and 1 year of scan editing experience, respectively. The number of scans that Observers 2–4 had edited before this study ranged from 10 to 100; Observer 1 has collected, edited, and analyzed over 1000 scans of some skeletal element.

An undergraduate student with limited knowledge in age estimation using skeletal remains was introduced to the study. As Observer 5, s/he was blinded to the study's purpose. After being provided with only minimal instruction on scan editing via online texts, Observer 5 edited the three sets of the raw SB scans at the same pace as Observers 2–4.

The raw SB scans were edited using the scanner's accompanying software, *ScanStudio 2.0.2*, such that the symphyseal face is extracted from the surrounding bones. From these isolated faces, x , y , and z coordinates were retrieved and analyzed via the *SAH* Score method (39), the TPS/BE method (40), and the VC method (41) to compute shape measures. More detailed, step-by-step procedures on how the shape measures are calculated can be found in Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41). Lastly, these measures were subjected to five sets of regression models to obtain final age estimates for each replicate scan per observer: three simple linear regression model using BE values, VC values, or *SAH* scores, independently, and two multiple regression models that combine BE values and VC values or the *SAH* scores and VC values. Both the shape measures and final age estimates reported in this study were calculated via the software, *forAge*, developed by Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41).

Intra/Interobserver Error Test

Using the shape measures and final age estimates, a series of the intraclass correlation coefficients (ICCs) were calculated to evaluate intra- and interobserver reliability in scan editing. By definition, *interobserver* reliability indicates *the variation between two or more observers* who take measurements from the same set of sample subjects. *Intraobserver* reliability, on the other hand, is about *the variation between the measurements themselves when a single observer repeatedly measures the same sample* (50). ICC, a modification of Pearson's correlation coefficient (45), is one of the most commonly used statistics that assesses reproducibility of quantitative measurements when the measurements are taken by multiple observers (or multiple times by one observer over time) on the same sample using the same method. It evaluates congruency between the multiple measurements/observers via the degree of correlation and agreement. Such conformity is represented by a coefficient ranging 0–1; any values close to 1 indicate higher reliability. As there are six to ten different forms of ICC with different calculations based on the definition one follows (46,47), deciding what ICC to use should depend on the nature of the study and the type of agreement the observer wishes to quantify (48). We use the following four questions as general guidelines to help identify which ICC form to use (49,50): (i) Does/do the observer(s) evaluate every case in the same sample (one-way vs. two-way model); (ii) do the observers represent researchers at large or a specific group of researchers (random- vs. mixed-effect model); (iii) are the observers interested in comparing single scores/measures or average of multiple measures; (iv) is the research investigating general consistency or absolute agreement among scores generated by different observers/at different times?

For this study, the exact same set of scans were edited by all five observers and we were interested in evaluating individual scan-to-scan agreement between the randomly selected observers. Therefore, the two-way random ICC model, which calculates absolute agreement on single measures (e.g., each of the raw shape measure values and age estimates derived from the five observers), was selected. Theoretically, for a single observer study (i.e., *intraobserver* error test), it is recommended to use the mixed-effect ICC model as one observer's scores cannot be generalized to represent the entire population of researchers (47). However, there is no separate formula for the mixed-effect model for ICC, and the current computation only allows one to use the same formula for both mixed- and random-effect models,

but with slightly different interpretation. Therefore, we based our analysis of intraobserver error on the two-way random ICC model as well.

Our model chosen for ICC is based on ANOVA which is well known to be robust to moderate violations of normality (51,52). We chose this robust, parametric model-based ICC following the observer error test protocol that is most familiar to and recommended by the forensic anthropology literature (16,53).

Intraclass correlation coefficients was computed using the inter-rater reliability package, *irr* (54), in the statistical environment R (55). The package provides *F*-test statistics and 95% confidence intervals for the selected ICC model. Although there is no universally agreed definition on the acceptable ICC values, the often-cited guidelines suggested by Cicchetti and Sparrow (56) and Cicchetti (57) were used (Table 1). These guidelines are in line with those proposed by Landis and Koch (58) and Fleiss (59). Additionally, we took a more conservative approach on the ICC value interpretation by taking into consideration of both point ICC values and 95% confidence intervals (CIs). Here, the 95% CIs predict a probable range on which the “real” population ICC value may lie and they essentially measure how much statistically reliable our ICC point estimates are. For significance, we set $\alpha = 0.05$.

Additional Test on Different Editing Conditions

In addition to the test of the consistency in scan editing between observers, we simulated two extra editing situations where the practitioner misidentifies age-related traits due to unfamiliarity with the scan editing protocol and/or inexperience with skeletal biology. This specifically addresses two concerns regarding scan editing; (i) how far one should delete the margin around the symphyseal face for the scan to be useful for age estimation; and (ii) whether one should include or exclude the pubic tubercle when editing raw scans. Although the pubic tubercle is not necessarily part of the symphyseal face, we found Question 2 to be important to test because the tubercle is one of the age-informative traits that are often used for pubic symphysis age estimation, and it sometimes extends to the symphyseal face making it difficult to identify the exact face border when editing raw scans. Theoretically, either inclusion or exclusion of the pubic tubercle is expected to impact the VC values as the feature protrudes ventrally. In other words, the inclusion of the ventrally protruding tubercle may lead the method’s algorithms to falsely fit a differently sized circle along the ventral outline, further generating a different curvature value compared to the condition in which the tubercle is properly deleted (Fig. 1).

For these additional tests, Observer 3 chose a set of the SB casts and edited them with different widths of the margin (2 mm vs. 4 mm vs. 1 cm) left around the symphyseal face, and with/without the pubic tubercle. From these differently edited scans, final age estimates were attained. Possible effects of these conditions to the accuracy of the estimated ages were evaluated by comparing the estimated ages to documented chronological ages.

TABLE 1—The interpretation guidelines for intraclass correlation coefficients (ICC) by Cicchetti (57).

ICC value	Interpretation
<0.40	Poor reliability
0.40–0.59	Fair reliability
0.60–0.74	Good reliability
0.75–1.00	Excellent reliability

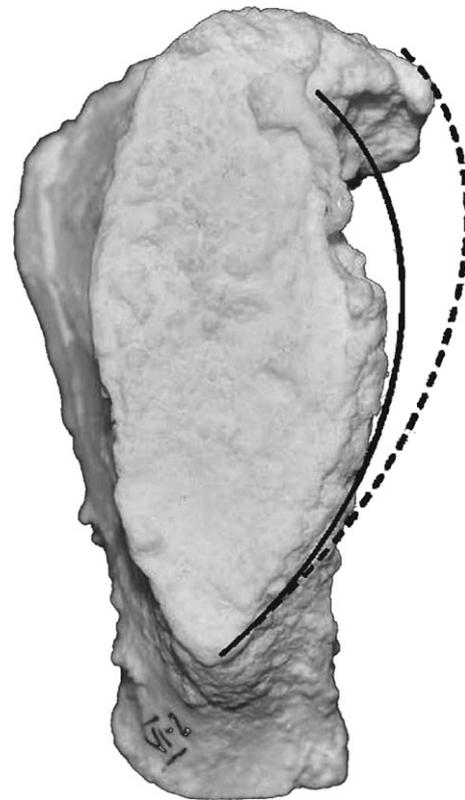


FIG. 1—A representation of two different ventral curvatures (solid and dashed lines) depending on the presence/absence of the tubercle (SB male cast upper phase VI).

To determine whether it is appropriate to use parametric statistical analyses with this dataset, the Shapiro–Wilk test of normality (*w*) was performed. The pairwise *t*-test was conducted to evaluate any difference in the mean age estimates and documented known ages. In addition, Pearson’s correlation coefficients (*r*) were calculated to assess a linear relationship between the age estimates of various editing conditions and documented known ages. The differences between age estimates and known ages were also visually evaluated using Bland–Altman plots. These analyses were conducted in the statistical program R (51) and/or JMP (60).

Results

Intraobserver Error

The majority of ICC values computed for the intraobserver error test fell between 0.75 and 1.00, indicating excellent reliability (Table 2). This supports that each of the four observers edited the raw SB scans consistently over three trials, generating compatible shape measures and age estimates to each other. The ICC of BE values for Observer 1 was slightly lower (0.728), but it did not have significant influence on final age estimates and still produced highly reliable BE age estimates for the three consecutive trials by the observer (0.910). The 95% CIs of the three shape measures (BE, *SAH* and VC values) showed more fluctuations in the range of reliability, between fair and excellent among the three trials per observer. However, the 95% CIs of the final age estimates generated from these shape measures, with only slightly varying values between trials, consistently showed excellent reliability, with an exception of Observer 4,

TABLE 2—Intraobserver error: ICC absolute agreement of within-observer shape measures and age estimates derived from first, second, and third time editing.

	ICC	DF (1,2)	F-Test	p-Value	Lower 95%	Upper 95%
Observer 1						
BE value	0.728	11,22.4	8.42	<0.001*	0.441	0.903
SAH score	0.856	11,22.6	20.5	<0.001*	0.675	0.952
VC value	0.942	11,23.9	48.8	<0.001*	0.859	0.981
BE estimate	0.910	11,23.1	29.7	<0.001*	0.785	0.971
SAH estimate	0.930	11,20	47.5	<0.001*	0.825	0.978
VC estimate	0.942	11,23.8	49	<0.001*	0.859	0.981
VC + BE estimate	0.949	11,24	56.4	<0.001*	0.875	0.984
VC + SAH estimate	0.945	11,20.4	60.3	<0.001*	0.861	0.982
Observer 2						
BE value	0.809	11,22.4	13.8	<0.001*	0.589	0.934
SAH score	0.780	11,22.6	11.3	<0.001*	0.536	0.923
VC value	0.924	11,23.9	35.8	<0.001*	0.817	0.975
BE estimate	0.920	11,23.9	34.7	<0.001*	0.809	0.974
SAH estimate	0.905	11,22.7	27.6	<0.001*	0.772	0.969
VC estimate	0.946	11,23.9	52.9	<0.001*	0.868	0.983
VC + BE estimate	0.952	11,22.9	56.5	<0.001*	0.880	0.985
VC + SAH estimate	0.940	11,23.6	46.3	<0.001*	0.854	0.981
Observer 3						
BE value	0.941	11,23.4	46.5	<0.001*	0.855	0.981
SAH score	0.913	11,24	32.7	<0.001*	0.793	0.971
VC value	0.933	11,22	39.6	<0.001*	0.835	0.979
BE estimate	0.953	11,23.9	61.2	<0.001*	0.885	0.985
SAH estimate	0.957	11,23.6	71	<0.001*	0.894	0.986
VC estimate	0.943	11,22.2	46.8	<0.001*	0.858	0.982
VC + BE estimate	0.972	11,22.6	97.4	<0.001*	0.928	0.991
VC + SAH estimate	0.969	11,23.8	98.6	<0.001*	0.923	0.990
Observer 4						
BE value	0.898	11,24	27.1	<0.001*	0.761	0.966
SAH score	0.978	11,21.8	152	<0.001*	0.944	0.993
VC value	0.828	11,23.1	14.7	<0.001*	0.620	0.942
BE estimate	0.817	11,23.4	13.8	<0.001*	0.599	0.937
SAH estimate	0.975	11,22.1	131	<0.001*	0.937	0.992
VC estimate	0.777	11,22.9	10.8	<0.001*	0.526	0.922
VC + BE estimate	0.789	11,22.4	11.4	<0.001*	0.545	0.927
VC + SAH estimate	0.973	11,23.9	112	<0.001*	0.933	0.991

*Significance level of $p < 0.05$

whose CIs of the age estimates ranged from fair to excellent (Table 2).

Interobserver Error

For the evaluation of four-observer error, a preliminary analysis was conducted to compute ICCs using both single measures ($n = 36$) and average of three trials per SB cast ($n = 12$). The resulting ICCs using average values were much more favorable as they were higher overall (ICCs > 0.9, See Table S1), compared to ICCs based on the single measures, which is due to the different calculations used for the average and single-measure approaches (see Field (49); Koo and Li (50) for the complete summary of ICC model formulae). Additionally, another set of ICCs were computed using each of the three trials per SB cast generated by the four observers using the two-way random ICC model on single measures for absolute agreement. While this approach reduces the total sample size to twelve instead of 36

TABLE 3—Interobserver error: ICC absolute agreement of between-observer shape measures and age estimates.

Shape measures/ Age estimates	ICC	DF (1,2)	F-Test	p-Value	Lower 95%	Upper 95%
BE value	0.865	35,108	26.8	<0.001*	0.790	0.921
SAH score	0.832	35,27.8	29.3	<0.001*	0.696	0.911
VC value	0.756	35,32.8	18	<0.001*	0.594	0.863
BE estimate	0.829	35,60.4	24.1	<0.001*	0.726	0.902
SAH estimate	0.836	35,32.6	28.7	<0.001*	0.711	0.911
VC estimate	0.746	35,28.7	17.8	<0.001*	0.571	0.859
BE + VC estimate	0.811	35,25	26.2	<0.001*	0.656	0.900
SAH + VC estimate	0.853	35,18.4	38.6	<0.001*	0.700	0.927

*Significance level of $p < 0.05$

TABLE 4—Intraobserver error of the inexperienced observer (Observer 5).

Shape Measures/ Age estimates	ICC	DF (1,2)	F-Test	p-Value	Lower 95%	Upper 95%
BE value	0.505	11,23.2	3.93	0.003*	0.150	0.800
SAH score	0.730	11,19.6	10.7	<0.001*	0.448	0.903
VC value	0.840	11,23.8	16.4	<0.001*	0.644	0.946
BE estimate	0.664	11,22.1	6.45	<0.001*	0.343	0.877
SAH estimate	0.753	11,21.6	11.4	<0.001*	0.490	0.912
VC estimate	0.828	11,24	15.5	<0.001*	0.624	0.941
VC + BE estimate	0.767	11,23.5	10.5	<0.001*	0.513	0.918
VC + SAH estimate	0.760	11,22.1	11.6	<0.001*	0.504	0.915

*Significance level of $p < 0.05$

scans, it allows us to compare, for example, Observer 1’s Trial 1, Observer 2’s Trial 1, . . . , Observer 4’s Trial 1, individually. Of 24 ICC values, the majority (20 ICC values) fell within the excellent reliability range (See Table S2), except four instances where ICCs were either on the borderline between good and excellent or in good agreement. However, we took a very conservative approach to the analysis in this study to ensure that we obtained cautioned estimates. Therefore, we report here the results of the single-measure absolute agreement ICCs using all 36 scans.

Similar to the results of the intraobserver error test, ICC values evaluating the four-observer error fell within the excellent reliability range (0.75–1.00) (Table 3). The point ICC values showed that, for the three shape measures, BE values had the highest reliability (0.865) among the four observers followed by the SAH scores (0.832) and VC values (0.756). The trend was consistent for final age estimates of the shape measures, although ICC of SAH score-based age estimates was slightly higher (0.836), while in a negligible degree, than that of the BE method (0.829). Of the two multiple regression models, age estimates of the VC + SAH method were most congruent between the four observers (0.853). VC values and VC age estimates were at the borderline between good and excellent reliability (0.756 and 0.746, respectively), indicating relatively greater observer variability in editing the ventral margin than the other two shape measures. This tendency is also shown in the 95% CIs: while most of the 95% CIs ranged from good to excellent reliability, the CIs of VC values and VC age estimates were wider ranging from fair to excellent reliability (Table 3). In addition, when the same analysis was run for only Observer 3 and Observer 4, who were not the developers of the methods, ICC

TABLE 5—Interobserver error after adding Observer 5.

Shape Measures/ Age Estimates	ICC	DF (1,2)	F-Test	p-Value	Lower 95%	Upper 95%
BE value	0.747	35,142	16.1	<0.001*	0.635	0.844
SAH score	0.593	35,11.6	19.4	<0.001*	0.303	0.780
VC value	0.665	35,23	17.9	<0.001*	0.456	0.810
BE estimate	0.744	35,50.3	20.2	<0.001*	0.604	0.849
SAH estimate	0.691	35,14.9	24.3	<0.001*	0.442	0.837
VC estimate	0.659	35,21.7	17.8	<0.001*	0.443	0.806
BE + VC estimate	0.712	35,19.4	23.4	<0.001*	0.499	0.844
SAH + VC estimate	0.710	35,12.3	29.7	<0.001*	0.442	0.853

*Significance level of $p < 0.05$

values were slightly lower than the analysis with the two developers included: BE value (0.769, excellent), followed by SAH + VC estimate (0.708, good), BE estimate (0.706, good), SAH score (0.688, good), SAH estimate (0.672, good), BE + VC estimate (0.656, good), VC value (0.632, good), VC estimate (0.599, between fair and good). The ICC values based on the average of the three trials for Observer 3 and Observer 4 were all in excellent agreement (see Table S3).

A Blind Introduction of an Inexperienced Observer

Intraclass correlation coefficients values of the inexperienced observer (Observer 5), who was blindly introduced to this study, ranged from fair to excellent reliability (0.505–0.840), indicating the observer edited the scans consistently throughout the three consecutive trials as the rest of the four observers (Table 4). Specifically, ICC of VC values and VC estimates was the highest (0.840 and 0.828, respectively) with the CIs ranging good to excellent reliability. However, the 95% CIs of Observer 5 were much wider compared to those of the previous four observers.

The widest 95% CI of Observer 5 was BE values, which ranged from 0.150 to 0.800, indicating poor to excellent reliability. We also conducted a preliminary analysis that considered the average of the three trials and the trial-by-trial comparisons, just as we explained for the four-observer tests above. We provide these preliminary results in Tables S4–S5.

In general, when the inexperienced observer was included to assess between-observer variation in scan editing, ICC values were slightly lower, and yet still the majority fell within the good reliability range (0.593–0.747) (Table 5). The highest ICC was BE values (0.747). The corresponding BE age estimates (0.744) were at the borderline between good and excellent reliability. The next two reliable shape measures were VC values (0.665) and SAH scores (0.593, fair reliability). Although the ICC of SAH scores was the lowest, suggesting that different scores were generated among observers, final age estimates produced using these variable SAH scores were still consistent between observers. The ICC for SAH score-based age estimates among the observers still showed good reliability (0.691). The lowest ICC value for age estimates was VC age estimates (0.659). The age estimates of the two multiple models had the same level of between-observer reliability (0.712, good reliability).

With the inclusion of Observer 5, the 95% CIs tend to fluctuate more and are much wider than those of the first four observers as seen in the Observer 5's intraobserver error results. Particularly, the 95% CI of SAH scores was 0.303–0.780 (poor-excellent reliability). This trend was absent when the observer error was assessed based on the four observers (the previous 95% CI of SAH scores was 0.696–0.911, good-excellent reliability), indicating the larger variability in the edited scans was coming from the scans of the inexperienced observer.

To summarize the five-observer error test, the ICCs of BE values and BE age estimates were the highest and deemed to be the most reliable between observers. This is consistent with the results of the four-observer analysis. As corroboration, we present Fig. 2, which compares the average BE value of the three trials generated

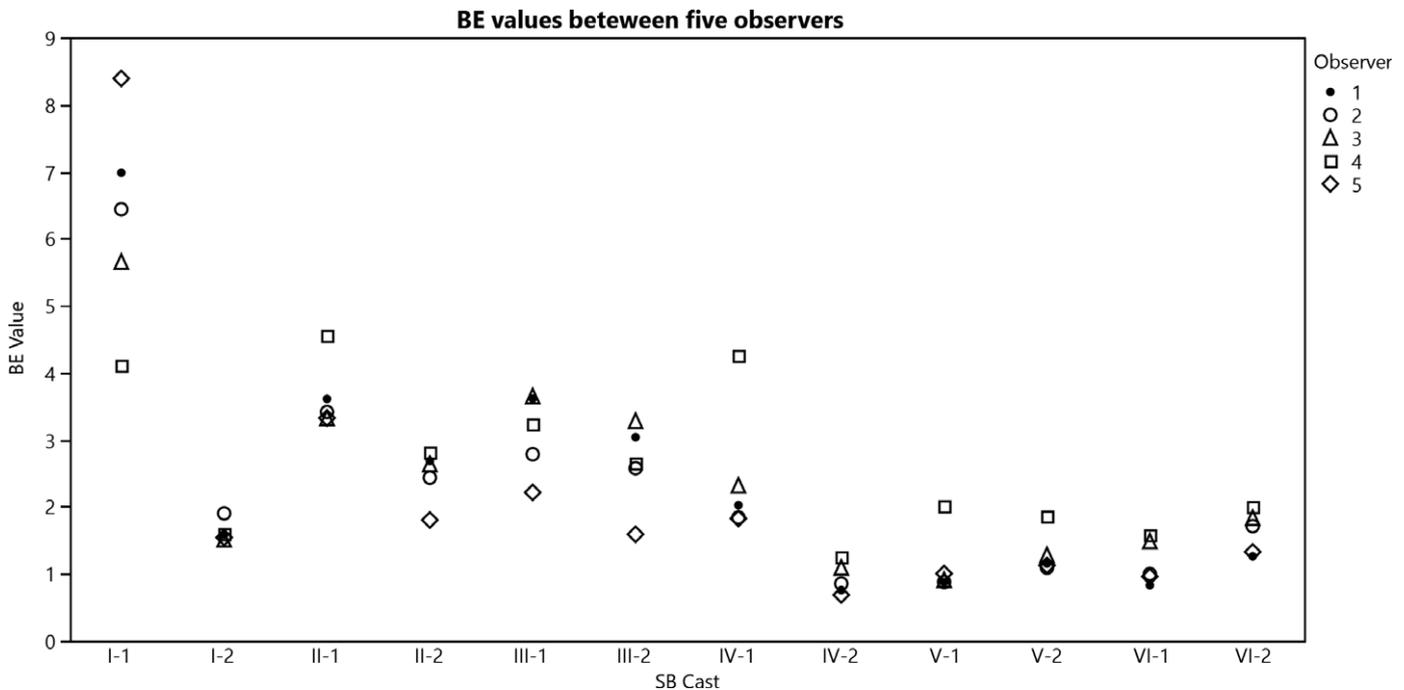


FIG. 2—BE value comparisons between observers. Note the proximity of BE values generated for each cast by five observers. An average BE value of the three trials was used for each observer.

per SB scan among the five observers. The proximity of the five BE values indicates the observers consistently and reliably edited the scans. Although the differences in BE values between observers were more prominent for the SB cast lower phase I (see Fig. 2), final age estimates generated from the various BE values were almost identical (see Fig. 3). However, this method tends to underestimate individuals in the upper SB phases (or older

individuals), as indicated by the age estimates that are located far below the known chronological ages of the SB casts (the asterisk symbol in Fig. 3). Nevertheless, we observed that the magnitude of underestimation reduces significantly when the multiple regression models were used to obtain final age estimates for the upper phases. Underestimation does persist, however, in two Phase VI casts, as shown in Fig. 4, which presents age estimates of the

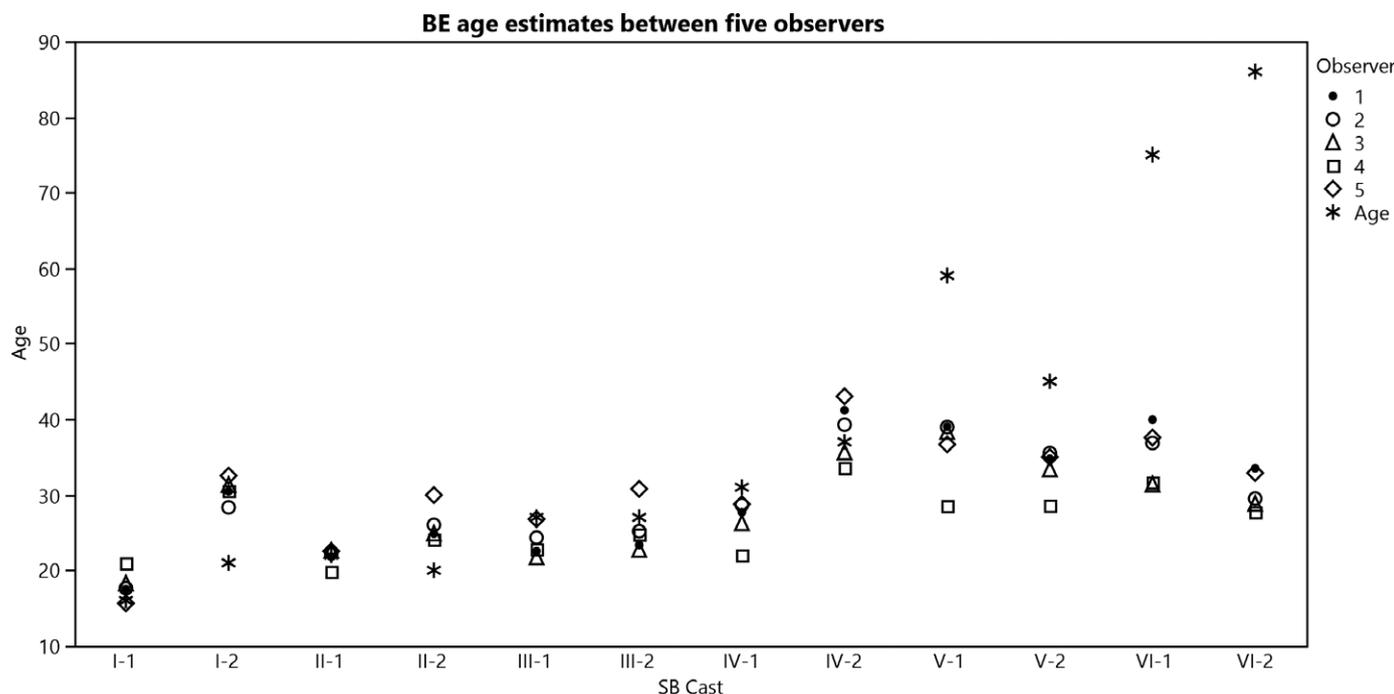


FIG. 3—Between-observer comparisons of age estimates generated from the BE values presented in Fig. 2. Note the overlapping age estimates among five observers. Known chronological ages of the SB casts are indicated by the asterisk symbol. An average BE age estimate of the three trials was used for each observer.

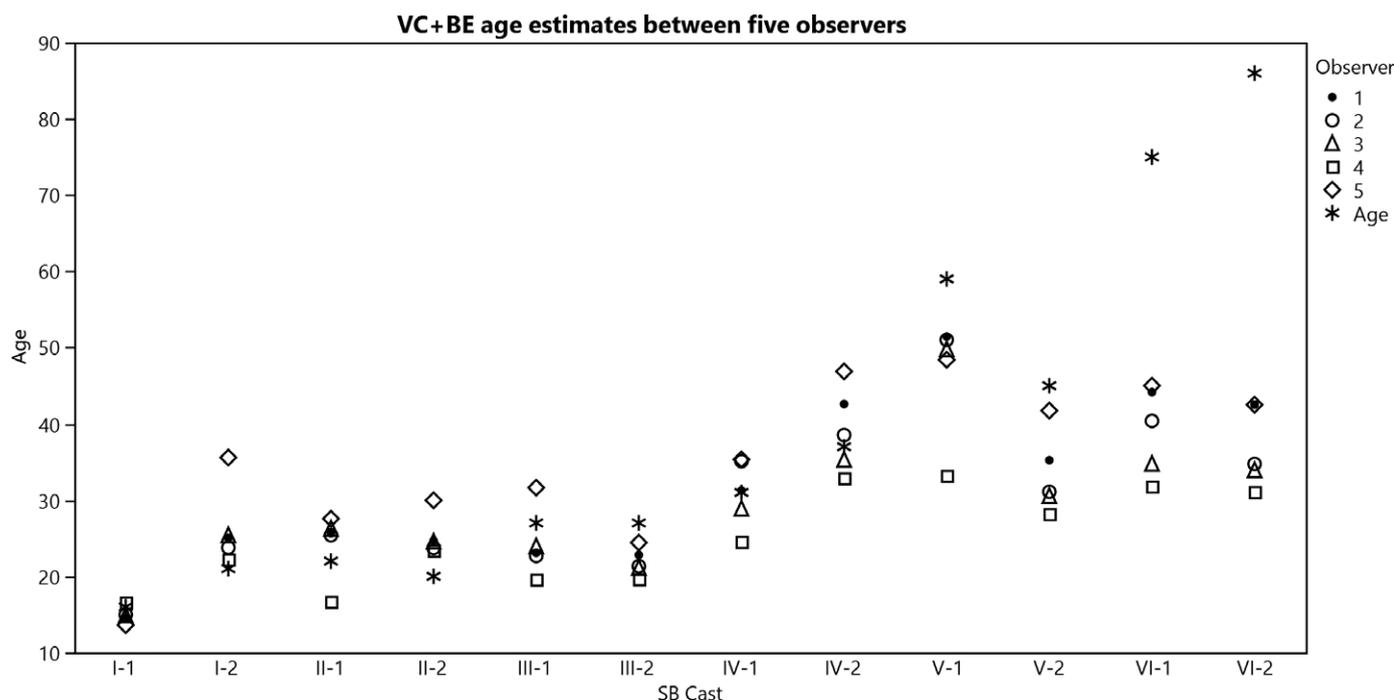


FIG. 4—Between-observer comparisons of age estimates generated from the multiple regression model that combines VC and BE values. Note the reduced magnitude of underestimation for the upper phase casts in comparison with the same casts in Fig. 3. Known chronological ages for the SB casts are indicated by the asterisk symbol. An average age estimate of the three trials was used for each observer.

regression model that combines BE and VC values. Based on these results, we can infer that BE method is robust to observer variability in scan editing, as it generates consistent shape measures and age estimates between different observers. The high reliability/repeatability of the BE method is further supported by its 95% CIs, which are relatively narrow, ranging good to excellent reliability, when compared to the rest of the CIs (Table 5).

Testing Different Editing Conditions

Effects of Different Symphyseal Margin Widths to Age Estimation—Results of the Shapiro–Wilk normality test indicated a normal distribution for most of the variables used in this study. Three variables, including the chronological ages, and the “W/ Tubercle” age estimates of the SAH method and the VC + SAH method, slightly deviated from the normality ($p = 0.037, 0.016, 0.048$, respectively Table 6). However, given that the t -test and

TABLE 6—The Shapiro–Wilk test of normality for all variables used in this study.

Shape Measure	Age Estimate per Condition	W	p-Value
BE	Age*	0.850	0.037 [†]
	2 mm	0.944	0.553
	4 mm	0.984	0.995
	W/ tubercle	0.883	0.097
	W/O tubercle	0.979	0.978
SAH	2 mm	0.945	0.567
	4 mm	0.918	0.268
	W/ tubercle	0.819	0.016 [†]
VC	W/O tubercle	0.888	0.111
	2 mm	0.958	0.759
	4 mm	0.897	0.145
VC + BE	W/ tubercle	0.980	0.964
	W/O tubercle	0.963	0.825
	2 mm	0.988	0.999
VC + SAH	4 mm	0.923	0.315
	W/ tubercle	0.923	0.347
	W/O tubercle	0.913	0.236
VC + SAH	2 mm	0.929	0.374
	4 mm	0.947	0.593
	W/ tubercle	0.854	0.048 [†]
	W/O tubercle	0.893	0.130

*Documented chronological ages

[†]Significance level of $p < 0.05$

Pearson’s correlation coefficients are known to be robust to even moderate deviations from normality (51,52) and that the non-parametric paired difference test, the Wilcoxon signed-rank test, yielded equally nonsignificant results, we report the results of the parametric tests here (the results of the nonparametric tests are presented in Tables S6–S7). Results of the pairwise t -test indicated that there was no significant difference between the means of the known ages and estimated ages derived from the pubic symphysis with either 2 or 4 mm margin left (Table 7). However, the mean of age estimates generated from the symphyseal face with 1 cm margin was significantly different from the mean of the documented ages, except for the VC methods (Table 7). This indicates that the 1 cm margin width was the threshold where the algorithms start generating statistically significantly different age estimates, especially for the two methods of SAH scores and BE values. Nonetheless, the computational methods mostly did not suffer from the improperly cleaned margins around the symphyseal face. In particular, the VC method excelled in generating correct age estimates despite the wide range of error (e.g., different margin widths, even with 1 cm margin) in scan editing.

Inclusion/Exclusion of the Pubic Tubercle—In contrast to our hypothesis, the VC method did not generate a statistically significant difference in mean estimated age and mean chronological age when the pubic tubercle was not removed from the raw scan. Instead, the shape-based methods evaluating the complexity of the pubic symphysis topography (i.e., BE and SAH scores) were more directly affected by the inclusion/exclusion of the pubic tubercle (Table 8). It is possible that the VC values were not affected by the inclusion of this extra feature because the tubercle is located in a different plane: the tubercle is slightly posterior to the symphyseal face surface when the observer is directly facing the surface. However, as the SAH score and BE methods measure the “ups and downs” of the symphyseal surface, those shape measures and age estimates are affected by the presence of the tubercle as the posteriorly extending feature adds another depth (or complexity) to the scan that the SAH scores and BE must account for. Accordingly, the scans with the tubercle tended to exacerbate underestimation for BE and SAH methods compared to the age estimates based on the scans without the tubercle. This trend is shown in the magnitude of mean

TABLE 7—Paired t -tests for mean comparisons between documented chronological ages and age estimates of the scans with three different margin widths.

Method	N	Mean Age	Mean age Estimate	Mean Difference	SE	Lower 95%	Upper 95%	t	DF	p-Value
2 mm										
BE	12	38.833	28.372	10.462	5.898	-2.520	23.443	1.774	11	0.104
SAH	12	38.833	34.765	4.068	4.003	-4.742	12.878	1.016	11	0.331
VC	11	40.454	39.820	0.634	5.712	-12.094	13.363	0.111	10	0.914
VC + BE	11	40.454	30.452	10.003	5.530	-2.319	22.324	1.809	10	0.100
VC + SAH	11	40.454	36.204	4.25	3.994	-4.647	13.149	1.064	10	0.312
4 mm										
BE	12	38.833	27.085	11.748	6.049	-1.566	25.063	1.942	11	0.078
SAH	12	38.833	31.786	7.047	4.587	-3.049	17.144	1.536	11	0.153
VC	11	40.454	39.373	1.082	6.057	-12.414	14.577	0.179	10	0.862
VC + BE	11	40.454	28.907	11.547	5.869	1.531	24.625	1.967	10	0.077
VC + SAH	11	40.454	33.042	7.413	4.663	-2.977	17.803	1.589	10	0.143
1 cm										
BE	12	38.833	24.570	14.263	6.023	1.005	27.522	2.368	11	0.037*
SAH	12	38.833	22.684	16.149	5.476	4.095	28.203	2.949	11	0.013*
VC	11	40.454	31.410	9.044	6.986	-6.520	24.609	1.294	10	0.224
VC + BE	11	40.454	22.935	17.519	6.595	2.823	32.215	2.656	10	0.024*
VC + SAH	11	40.454	22.483	17.972	6.066	4.455	31.489	2.962	10	0.014*

*Significance level of $p < 0.05$

TABLE 8—Paired *t*-tests for mean comparisons between documented chronological ages and age estimates of the scans with/without the tubercle.

Method	<i>N</i>	Mean Age	Mean age Estimate	Mean Difference	SE	Lower 95%	Upper 95%	<i>t</i>	DF	<i>p</i> -Value
With Tubercle										
BE	12	38.833	21.653	17.181	5.281	5.557	28.804	3.253	11	0.008*
SAH	12	38.833	23.140	15.693	3.804	7.321	24.066	4.125	11	0.002*
VC	11	40.455	36.485	3.969	6.390	-10.268	18.207	0.621	10	0.548
VC + BE	11	40.455	23.052	17.403	5.366	5.447	29.358	3.243	10	0.009*
VC + SAH	11	40.455	24.310	16.145	3.941	7.364	24.925	4.099	10	0.002*
Without Tubercle										
BE	12	38.833	27.343	11.491	5.873	-1.436	24.417	1.956	11	0.076
SAH	12	38.833	33.070	5.763	4.114	-3.291	14.818	1.401	11	0.189
VC	11	40.455	38.925	1.529	5.925	-11.673	14.732	0.258	10	0.801
VC + BE	11	40.455	29.008	11.446	5.682	-1.213	24.106	2.015	10	0.072
VC + SAH	11	40.455	34.323	6.132	4.266	-3.374	15.638	1.437	10	0.181

*Significance level of $p < 0.05$

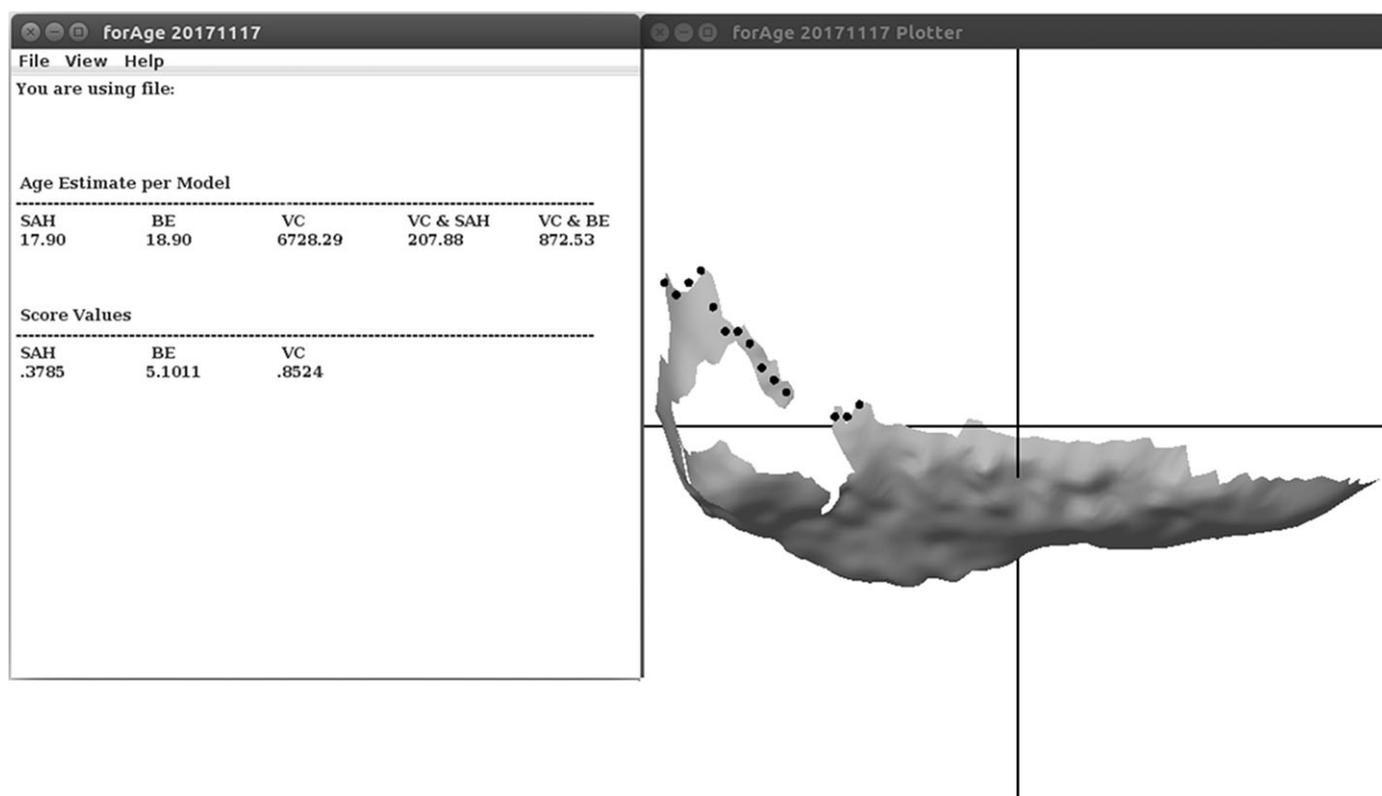


FIG. 5—An example of misloading of the edited scan by *forAge*, when the symphyseal face scan includes the pubic tubercle. Note the very unrealistic age estimates for SB male cast upper phase I.

differences between two different editing conditions in Table 8. The mean age estimates of the multiple regression models were also significantly different from the mean documented age as the models include the SAH scores and BE/TPS values (VC + BE $t(10) = 3.243$, $p = 0.009$; VC + SAH $t(10) = 4.099$, $p = 0.002$, Table 8).

In addition, when the edited scans of the symphyseal face included the pubic tubercle or had too much of the margins left (e.g., 1 cm), *forAge* loaded the scans in the wrong direction with incorrect age estimates, sometimes generating very unlikely to impossible age estimates (e.g., 6728.29 years, Fig. 5). When this happens, we advise the practitioner to cease the analysis and re-edit the scan. For this study, we observed three instances where the scans with the tubercle yielded unrealistic age estimates: a single variable VC model,

VC + SAH model, and VC + BE model (6728.29, 872.53, 207.88 years respectively) for the SB cast Phase I-2. To facilitate the statistical analysis, these numbers were treated as missing values.

As expected, when the pubic tubercle is properly removed from the scan, there was no significant difference between the means of the estimated and documented ages (Table 8). In support of these results, the Pearson correlation coefficients suggested there was a positive linear relationship between the known ages and age estimates of various editing conditions ($r = 0.449-0.931$, Table 9), indicating that both the chronological age and estimated age increase in the same direction. It is important to note, however, that the correlations for the BE-based method did not differ significantly from zero ($\alpha = 0.05$): especially for three editing conditions with 2 mm

TABLE 9—Pearson's correlation coefficients between documented chronological ages and age estimates under various editing conditions.

Method	Age Estimate	<i>r</i>	<i>N</i>	Lower 95%	Upper 95%	<i>p</i> -Value
BE	2 mm	0.503	12	-0.100	0.836	0.096
	4 mm	0.463	12	-0.151	0.819	0.130
	W/ Tubercle	0.787	12	0.388	0.937	0.002*
SAH	W/O Tubercle	0.555	12	-0.027	0.856	0.061
	2 mm	0.879	12	0.616	0.966	<0.001*
	4 mm	0.806	12	0.433	0.944	0.002*
VC	W/ Tubercle	0.904	12	0.687	0.973	<0.001*
	W/O Tubercle	0.893	12	0.654	0.970	<0.001*
	2 mm	0.639	12	0.104	0.888	0.025*
VC + BE	4 mm	0.616	12	0.065	0.879	0.033*
	W/ Tubercle	0.449	11	-0.206	0.826	0.166
	W/O Tubercle	0.604	12	0.046	0.875	0.038*
VC + SAH	2 mm	0.677	12	0.169	0.901	0.016*
	4 mm	0.655	12	0.130	0.893	0.021*
	W/ Tubercle	0.830	11	0.458	0.955	0.002*
VC + SAH	W/O Tubercle	0.649	12	0.120	0.891	0.022*
	2 mm	0.876	12	0.607	0.965	<0.001*
	4 mm	0.834	12	0.499	0.952	0.001*
VC + SAH	W/ Tubercle	0.931	11	0.751	0.982	<0.001*
	W/O Tubercle	0.867	12	0.584	0.962	<0.001*

*Significance level of *p* < 0.05

margin, with 4 mm margin, and without the tubercle. The VC method's age estimates for scans with the tubercle also showed nonsignificant correlations with the chronological ages. This lack of a significance between VC/BE age estimates and chronological ages may be attributable to tendency for underestimation by the univariate regression models, as already discussed previously in relation to Figs 2–4, rather than to observer-induced scan editing error.

Finally, the Bland–Altman plots in Fig. 6 demonstrate how the computational methods produced persistent underestimation regardless of the five different editing conditions, especially for the upper phase casts: on average as small as 4 years and as large as 17 years for the SAH score, BE, and two multiple regression methods. The VC method showed the least underestimation, as the mean difference between the estimated and known ages ranged from 0.6 to 9 years (see Fig. 6 and Tables 7–8).

Discussion

In response to our stated concerns over method reliability in the presence of observer variation, the current paper evaluates the repeatability of the new fully computational methods proposed by Slice and Algee-Hewitt (39) and Stoyanova et al. (40,41), and, more broadly speaking, the degree of error associated with using high-dimensional laser scan data, when testing multiple, randomly selected observers with diverse training backgrounds and experience levels. In this study, we have demonstrated that consistent and comparable shape values and final age estimates were generated from scans edited by five independent observers with various experience levels and training backgrounds. Overall, the shape measures that evaluate the complexity in the symphyseal face topography (e.g., BE and SAH methods) are most resistant to observer error and scan editing variability. On the other hand, the VC method tends to be relatively more sensitive to such variation, although ICC was still within the range of excellent reliability for the four-observer study and good reliability for the five-observer study, which included an inexperienced observer. Thus, our results find that lower ICC values in three shape measures do not necessarily mean that the associated final age estimates will also be different between observers: the algorithms generate somewhat different shape measures in response to the scans edited by different observers, but the final age estimates are still similar. Moreover, none of our results for the multiple regression models present particularly low ICC, suggesting that the two shape measures in the models compensate for any possible observer variation in the edited scans and generate matched age estimates among observers. Lastly, we show that the computational methods are not as susceptible to various inappropriate editing conditions as we may have expected. All five age estimation models based on the three shape measures generate consistently accurate and reliable age estimates, except two conditions where the scans had the 1 cm margin left around the pubic symphysis and the pubic tubercle. Despite its relative sensitivity to observer variation, the VC method performed the best for all five different editing conditions. Although the inclusion of the pubic tubercle did not result in inaccurate age estimates for the VC method, we still recommend removing the tubercle as it may influence age estimates of the BE and SAH methods and the two multiple models.

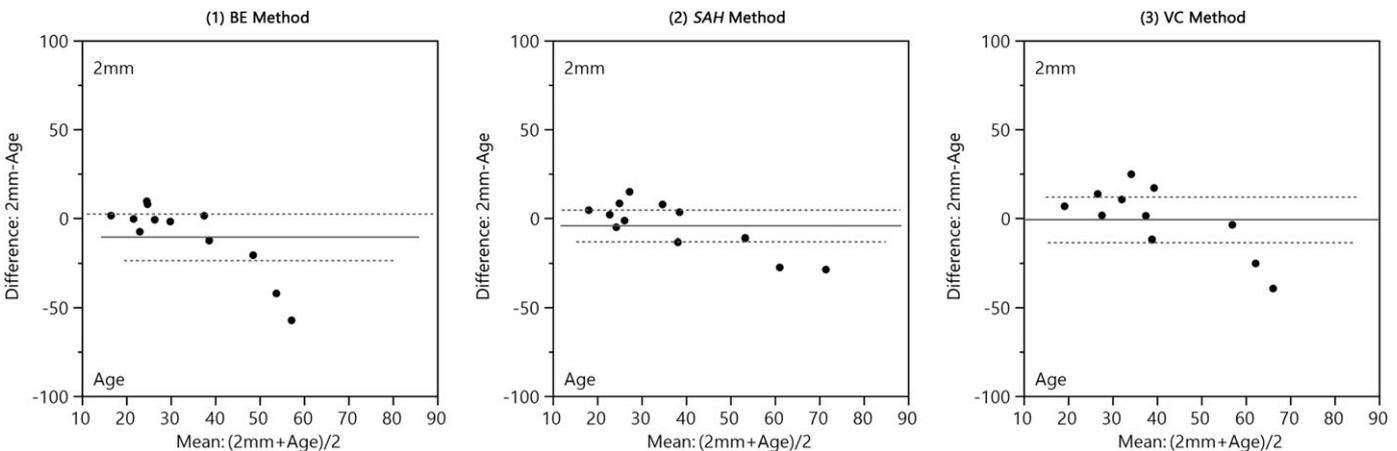


FIG. 6—Bland–Altman plots showing the difference between the documented chronological ages and the age estimates of the three shape measures based on the scans with 2 mm margin: (i) BE method, mean difference ≈ negative 10 years (*p* = 0.104); (ii) SAH score method, mean difference ≈ negative 4 years (*p* = 0.331); and (iii) VC method, mean difference ≈ negative 0.6 years (*p* = 0.914).

Similarly, the margins of the pubic symphysis need to be trimmed as close to the rim as possible as the *SAH* score and BE value should be computed based on the data points distributed within the symphyseal face, not beyond. In conclusion, this more complex (multiple inter-/intra-) observer error study corroborates the results of the single observer error study reported in Stoyanova et al. (40). We have confirmed the high repeatability in the shape-based age estimation based on 3D laser scans of the pubic symphysis. In doing so, this study indicates that these fully computational methods, and possibly more broadly, any future methods that rely similarly upon 3D scan editing, can be widely used with confidence, even by users with limited skeletal knowledge, age estimation experience, training in the laser scan editing techniques, and the implementation of the method-specific software, *forAge*. We conclude that the semi-automated framework of analysis that uses the computational methods proposed by Slice and Algee-Hewitt (39) and Stoyanova et al.(40,41) can indeed serve as a standardized procedure that enables modern skeletal researchers and forensic laboratories to produce highly compatible age-at-death estimates with a reduced risk of the observer effects that may otherwise contribute to inaccurate, unreliable, and/or biased age estimates.

Acknowledgments

The authors are grateful to Dr. Judy Suchey for providing the documented ages of the Suchey–Brooks male casts, and Maddison Woods at the Department of Anthropology, Florida State University, for editing the raw Suchey–Brooks cast scans for this project.

References

- Plato C, Fox K, Tobin J. Skeletal changes in human aging. In: Crews DE, Garruto RM, editors. *Biological anthropology and aging: perspectives on human variation over the life span*. New York, NY: Oxford University Press, 1994;272–300.
- Kemkes-Grottenthaler A. Aging through the ages: historical perspectives on age indicator methods. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Publishing, 2002;48–72.
- Kobyliansky E, Karasik D, Belkin V, Livshits G. Bone ageing: genetics versus environment. *Ann Hum Biol* 2000;27(5):433–51.
- Algee-Hewitt BFB. Age estimation in modern forensic anthropology: statistical considerations. In: Tersigni-Tarrant MA, Shirley NR, editors. *Forensic anthropology: a comprehensive introduction*. Boca Raton, FL: CRC Press, 2017;381–420.
- Milner GR, Wood JW, Boldsen JL. Advances in paleodemography. In: Katzenberg MA, Saunders SR, editors. *Biological anthropology of human skeleton*, 2nd edn. New York, NY: John Wiley and Sons Inc, 2008;561–600.
- Boldsen JL, Milner GR, Konigsberg LW, Wood JW. Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press, 2002;73–106.
- Buckberry JL, Chamberlain AT. Age estimation from the auricular surface of the ilium: a revised method. *Am J Phys Anthropol* 2002;119(3):231–9.
- Wilczak CA, Christopher DJ. *Osteoware software manual*. vol. I. Washington, DC: Smithsonian Institution, 2011.
- McKern TW, Stewart TD. Skeletal age changes in young American males: analyzed from the standpoint of age identification. Quartermaster Research and Development Command 764 Technical Report EP-45. Natick, MA: US Army Quartermaster Research and Development Center, 1957.
- Bocquet-Appel J-P, Masset C. Farewell to paleodemography. *J Hum Evol* 1982;11(4):321–33.
- Masset C. Age estimation based on cranial sutures. In: Iscan MS, editor. *Age markers in the human skeleton*. Springfield, IL: C.C. Thomas, 1989;71–103.
- Konigsberg LW, Frankenberg SR, Walker RB. Regress what on what? Paleodemographic age estimation as a calibration problem. In: Paine RR, editor. *Integrating archaeological demography: multidisciplinary approaches to prehistoric population*. Carbondale, IL: Southern Illinois University Press, 1997;24:64–88.
- Usher BM. Reference samples: the first step in linking biology and age in the human skeleton. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press, 2002;29–47.
- Komar DA, Grivas C. Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. *Am J Phys Anthropol* 2008;137(2):224–33.
- Kimmerle EH, Prince DA, Berg GE. Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth. *J Forensic Sci* 2008;53(3):594–600.
- Shirley NR, Ramirez Montes PA. Age estimation in forensic anthropology: quantification of observer error in phase versus component-based methods. *J Forensic Sci* 2015;60(1):107–11.
- Milner GR, Boldsen JL. Transition analysis age estimation: skeletal scoring manual. *Fordisc* Version 1.02, 2016.
- Rissech C, Estabrook GF, Cunha E, Malgosa A. Using the acetabulum to estimate age at death of adult males. *J Forensic Sci* 2006;51(2):213–29.
- Rissech C, Estabrook GF, Cunha E, Malgosa A. Estimation of age-at-death for adult males using the acetabulum, applied to four Western European populations. *J Forensic Sci* 2007;52(4):774–8.
- San-Millán M, Rissech C, Turbón D. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *Int J Legal Med* 2017;131(2):501–25.
- Hoppa R, Vaupel J. The Rostock Manifesto for paleodemography: the way from stage to age. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press, 2002;1–8.
- Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 2008;53(3):541–57.
- Konigsberg LW, Frankenberg SR. Bays in biological anthropology. *Am J Phys Anthropol* 2013;152(Suppl 57):153–84.
- Konigsberg LW, Frankenberg SR, Liversidge HM. Optimal trait scoring for age estimation. *Am J Phys Anthropol* 2016;159(4):557–76.
- Godde K, Hens SM. Modeling senescence changes of the pubic symphysis in historic Italian populations: a comparison of the Rostock and forensic approaches to aging using transition analysis. *Am J Phys Anthropol* 2015;156(3):466–73.
- Prince DA, Kimmerle EH, Konigsberg LW. A Bayesian approach to estimate skeletal age-at-death utilizing dental wear. *J Forensic Sci* 2008;53(3):588–93.
- Kimmerle EH, Konigsberg LW, Jantz RL, Baraybar JP. Analysis of age-at-death estimation through the use of pubic symphyseal data. *J Forensic Sci* 2008;53(3):558–68.
- Milner G, Boldsen J. Skeletal age estimation: where we are and where we should go. In: Dirksmaat DC, editor. *A companion to forensic anthropology*. West Sussex: Blackwell Publishing, 2012;224–38.
- Villa C, Hansen MN, Buckberry J, Cattaneo C, Lynnerup N. Forensic age estimation based on the trabecular bone changes of the pelvic bone using post-mortem CT. *Forensic Sci Int* 2013;233(1):393–402.
- Villa C, Buckberry J, Cattaneo C, Lynnerup N. Reliability of suchey-brooks and buckberry-chamberlain methods on 3D visualizations from CT and laser scans. *Am J Phys Anthropol* 2013;151(1):158–63.
- Villa C, Buckberry J, Cattaneo C, Frohlich B, Lynnerup N. 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* 2015;60(3):556–65.
- Boyd KL, Villa C, Lynnerup N. The use of CT scans in estimating age at death by examining the extent of ectocranial suture closure. *J Forensic Sci* 2015;60(2):363–9.
- Villa C, Buckberry J, Lynnerup N. Evaluating osteological ageing from digital data. *J Anat*. <https://doi.org/10.1111/joa.12544>. Epub 2016 Sep 13
- Navega D, Coelho JDO, Cunha E, Curate F. DXAGE: a new method for age at death estimation based on femoral bone mineral density and artificial neural networks. *J Forensic Sci* 2017;63(2):497–503.
- Curate F, Albuquerque A, Cunha EM. Age at death estimation using bone densitometry: testing the Fernandez Castillo and Lopez Ruiz method in two documented skeletal samples from Portugal. *Forensic Sci Int* 2013;226(1-3):296.e1–6.

36. Lottering N, Macgregor DM, Meredith M, Alston CL, Gregory LS. 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* 2013;150(3):386–99.
37. Lottering N, Reynolds MS, MacGregor DM, Meredith M, Gregory LS. Morphometric modelling of ageing in the human pubic symphysis: sexual dimorphism in an Australian population. *Forensic Sci Int* 2014;236(195):e1–11.
38. Chiba F, Makino Y, Motomura A, Inokuchi G, Torimitsu S, Ishii N, et al. Age estimation by quantitative features of pubic symphysis using multidetector computed tomography. *Int J Legal Med* 2014;128(4):667–73.
39. Slice DE, Algee-Hewitt BF. Modeling bone surface morphology: a fully quantitative method for age-at-death estimation using the pubic symphysis. *J Forensic Sci* 2015;60(4):835–43.
40. Stoyanova D, Algee-Hewitt BF, Slice DE. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am J Phys Anthropol* 2015;158(3):431–40.
41. Stoyanova DK, Algee-Hewitt BF, Kim J, Slice DE. 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* 2017;62(6):1434–44.
42. Figueroa-Soto C, Stoyanova D, Kim J, Slice DE, Algee-Hewitt BFB. Testing computational age estimation methods using laser scans of the adult pubic symphysis on modern Hispanic populations. Proceedings of the 69th Annual Scientific Meeting of the American Academy of Forensic Sciences; 2017 Feb 13-18; New Orleans, LA. Colorado Springs, CO: American Academy of Forensic Sciences, 2017:122.
43. Stoyanova D, Kim J, Figueroa-Soto C, Slice DE, Algee-Hewitt BFB. Age-at-death estimation based on the female pubic symphysis using computational methods and 3D laser scans. *Am J Phys Anthropol* 2018;165(S66):266.
44. Stoyanova D, Algee-Hewitt BFB, Kim J, Figueroa-Soto C, Slice DE. A study on the asymmetry between the left and right human pubic symphysis for age-at-death estimation based on 3D laser scans and computational methods. Proceedings of the 70th Annual Scientific Meeting of the American Academy of Forensic Sciences; 2018 Feb 19-24; Seattle, WA. Colorado Springs, CO: American Academy of Forensic Sciences, 2018:63.
45. Fisher RA. *Statistical methods for research workers*. New Delhi, Delhi: Genesis Publishing Pvt Ltd, 1925.
46. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1(1):30–46.
47. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–8.
48. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 2012;8(1):23–4.
49. Field AP. *Intraclass correlation*. Wiley StatsRef: Statistics Reference Online, 2005.
50. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155–63.
51. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health* 2002;23(1):151–69.
52. Quinn GP, Keough MJ. *Experimental design and data analysis for biologists*. Cambridge: Cambridge University Press, 2002.
53. Lottering N, MacGregor DM, Alston CL, Watson D, Gregory LS. Introducing computed tomography standards for age estimation of modern Australian subadults using postnatal ossification timings of select cranial and cervical sites. *J Forensic Sci* 2016;61(Suppl 1):S39–52.
54. Gamer M, Lemon J, Fellows I, Singh P. irr: Various coefficients of interrater reliability and agreement. R package Version 0.84, 2012. <https://rdrr.io/cran/irr/> (accessed June 5, 2018).
55. R Core Team. *A language and environment for statistical computing*. R. Vienna, Austria: R Foundation for Statistical Computing, 2017.
56. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86(2):127–37.
57. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6(4):284–90.
58. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–74.
59. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*, 3rd edn. Hoboken, NJ: John Wiley & Sons Inc, 2003.
60. Cary N. JMP®, Version 13. Cary, NC: SAS Institute Inc., 1989–2007.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1. Inter-observer error (four observers): ICC absolute agreement of between-observer shape measures and age estimates based on the average of three trials of each cast.

Table S2. Inter-observer error (four observers): ICC absolute agreement of between-observer shape measures and age estimates for each trials.

Table S3. Inter-observer error between Observer 3 and Observer 4: ICC absolute agreement of between-observer shape measures and age estimates.

Table S4. Inter-observer error (five observers): ICC absolute agreement of between-observer shape measures and age estimates based on the average of three trials of each cast.

Table S5. Inter-observer error (five observers): ICC absolute agreement of between-observer shape measures and age estimates for each trial.

Table S6. Non-parametric, matched-pair Wilcoxon signed rank tests for comparisons between documented chronological ages and age estimates of the scans with various editing conditions.

Table S7. Non-parametric Spearman's rank correlation coefficients (ρ) between documented chronological ages and age estimates under various editing conditions.

An Enhanced Computational Method for Age-at-Death Estimation Based on the Pubic Symphysis Using 3D Laser Scans and Thin Plate Splines

Detelina Stoyanova,¹ Bridget F.B. Algee-Hewitt,^{1,2,3*} and Dennis E. Slice^{1,4}

¹Department of Scientific Computing, Florida State University, Tallahassee, FL 32306

²Department of Biology, Rosenberg Lab, Stanford Center for Computational, Evolutionary and Human Genomics, Stanford University, Stanford, CA 94305

³Department of Anthropology, Florida State University, Tallahassee, FL 32306

⁴Department of Anthropology, University of Vienna, Vienna 1090, Austria

KEY WORDS biological profile; forensic case analysis; quantitative methods; paleodemography; bioarchaeology

ABSTRACT

OBJECTIVES: The pubic symphysis is frequently used to estimate age-at-death from the adult skeleton. Assessment methods require the visual comparison of the bone morphology against age-informative characteristics that represent a series of phases. Age-at-death is then estimated from the age-range previously associated with the chosen phase. While easily executed, the "morphoscopic" process of feature-scoring and bone-to-phase-matching is known to be subjective. Studies of method and practitioner error demonstrate a need for alternative tools to quantify age-progressive change in the pubic symphysis. This article proposes a more objective, quantitative method that analyzes three-dimensional (3D) surface scans of the pubic symphysis using a thin plate spline algorithm (TPS).

MATERIALS AND METHODS: This algorithm models the bending of a flat plane to approximately match the

surface of the bone and minimizes the bending energy required for this transformation. Known age-at-death and bending energy were used to construct a linear model to predict age from observed bending energy. This approach is tested with scans from 44 documented white male skeletons and 12 casts.

RESULTS: The results of the surface analysis show a significant association (regression p -value = 0.0002 and coefficient of determination = 0.2270) between the minimum bending energy and age-at-death, with a root mean square error of ≈ 19 years.

DISCUSSION: This TPS method yields estimates comparable to established methods but offers a fully integrated, objective and quantitative framework of analysis and has potential for use in archaeological and forensic casework. *Am J Phys Anthropol* 158:431–440, 2015. © 2015 Wiley Periodicals, Inc.

Physical anthropologists have long been interested in the morphological variation observed within and among skeletal populations, and there remains a longstanding desire to make use of these observations in order to reconstruct past population history, to understand the biological consequences of different life ways, and, most recently, to build individual biological profiles for the identification of unknown remains in medicolegal contexts. These ongoing efforts have resulted in an extensive and evolving toolkit that skeletal biologists can call upon to evaluate the physical characteristics of osteological remains and to estimate the biological parameters (like sex, age, stature, or ancestry) that are of interest to paleodemographic studies and useful for human identification casework. Accordingly, when the goal is to answer the question of age-at-death for the adult skeleton, the anthropologist looks to an established suite of age-progressive bony indicators, assessment techniques, and estimation methods that can be simply used with high repeatability and low error. Together these resources enable the translation of biological age information collected from the skeletal remains into a statistical estimate of lived years for the individual or population under study (Algee-Hewitt, 2013).

Standards of practice for estimating age-at-death for the adult skeleton typically require the application of component-scoring or phase-based aging methods to the

articular surfaces of the pelvis (e.g., McKern and Stewart, 1957; Lovejoy et al., 1985; Brooks and Suchey, 1990). The pubic symphysis, in particular, is said to be both the "most reliable" (Buikstra and Ubelaker, 1994: 21) and "most frequently used" (Meindl et al., 1985: 29) indicator for accessing information on age-at-death. It has been shown for past and present populations that throughout adult life the surface of the pubic symphysis undergoes post-developmental, degenerative modification at a regular and predictable rate. Given this age-progressive change, it should not be surprising that for many decades, researchers have been interested in developing techniques for age-estimation that capitalize upon the morphological variation in this structure. Todd (1920) first described the patterns of macroscopic change

*Correspondence to: Bridget F.B. Algee-Hewitt, Department of Biology, Rosenberg Lab, Gilbert Building, Room 109, 371 Serra Mall, Stanford University, Stanford, CA 94305-5020, USA. E-mail: bridgeta@stanford.edu

Received 10 February 2015; revised 8 June 2015; accepted 9 June 2015

DOI: 10.1002/ajpa.22797
Published online 14 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

in the pubic bones, established scoring criteria for capturing this age-progression in a suite of key traits under a multiphase classification system, and provided a definitive set of visual assessment standards. With the increasing recognition of the utility of Todd's approach, a number of researchers worked in the years that followed to provide more refined methods of age-at-death estimation that were based on Todd's original achievements (e.g., Brooks, 1955; McKern and Stewart, 1957).

The Suchey-Brooks system (Brooks and Suchey, 1990) has, thus far, made best use of the age progressive changes in adult symphyseal morphology and, so, it continues to remain the "most highly favored aging technique" for skeletal case analysis (Garvin and Passalacqua, 2012). The Suchey-Brooks method combined Todd's original ten phases into six phases of aging for each sex and provides written descriptions and illustrations for a visual assessment: the pubic symphysis is compared with the physical details that characterize each of the six phases, and it is assigned to the phase that best captures its state (i.e., degree, form) of skeletal expression (Brooks and Suchey, 1990). The six phases are characterized by the changes that occur on the shape of the symphyseal face. For adult individuals it starts as a set of high horizontal ridges and furrows: this billowing appears to fill-in as it flattens-out over time. The symphyseal face is usually smooth by phase 4 and after that it becomes pitted and porous. In addition, by phase 4 the symphyseal face starts to develop a distinct rim, during phase 5 the face is completely rimmed, and this rim may later erode (Brooks and Suchey, 1990). Despite the ease with which the Suchey-Brooks technique can be applied to the pubic symphysis and the arguments made for the regularity of the morphological changes that underlie its categorical framework (Suchey et al., 1986; Suchey and Katz, 1998), recent criticisms of this, as well as other phase-age and ordinal scoring systems, have identified several shortcomings (Baccino et al., 1999; Kimmerle et al., 2008; Berg, 2008; Hartnett, 2010). There is, for example, 1) a very large overlap among the wide age-ranges provided for each of the six phases, 2) a loss of precision when the technique is applied to older individuals (e.g., 40+ years), 3) a great waste of morphological information on age that is not captured by the feature-scoring and phase-assignment criteria, and 4) a high degree of variability in method interpretation and a low degree of standardization across applications and practitioners (Algee-Hewitt, 2013). These factors all imply that this current "morphoscopic" process of feature-scoring and bone-to-phase-matching allows for a very large degree of observation (user-induced) and estimation (method-specific) error. Assessment criteria do have the potential for improvement, and efforts are being made to develop more effective component-driven approaches that consider individually the sequence and the rate of change for the features contributing to the indicator's morphology (Milner and Boldsen, 2012, 2013). All of the adult age estimation techniques, whether based on the pubic symphysis or other age-progressive/skeletal degenerative indicators, that are qualitative in their methodologies, share the disadvantage of being dependent, to some nontrivial extent, on the expectations of the researchers and experience of the practitioners (Baccino et al., 1999; Kimmerle, Prince and Berg, 2008).

What is still needed to advance the science of age-at-death estimation is the widespread adoption of objective,

quantitative methods that better capture the full range of diversity in the morphological expression of the indicator at any given age, reduce the user-induced error by removing the subjectivity of the scoring procedure, and improve the estimation of point-ages and age ranges by introducing a statistically robust, fully computational workflow (Slice and Algee-Hewitt, 2015). Further, they should respond to the ongoing calls for implementing standardized, best-practice methods for age-at-death estimation (Garvin and Passalacqua, 2012). It is increasingly important that emerging methods yield the quality of skeletal aging data necessary to make meaningful paleodemographic inferences in archaeological research contexts and to meet the evolving medico-legal standards of scientific evidence for forensic casework reporting (Ritz-Timme et al., 2000; Hoppa and Vaupel, 2002; Christensen and Crowder, 2009; Holden, 2009; SWGANTH Age Estimation, 2013; Algee-Hewitt, 2013).

The purpose of this study is to develop an objective fully quantitative method for age-at-death determination. Here, we present a unique approach that exploits the potential of three-dimensional (3D) surface scans of the pubic symphysis, implements a thin plate spline (TPS) algorithm, and allows for a straightforward and consistent estimation in both archaeological and forensic settings.

MATERIAL AND METHODS

Surface scan files

In order to analyze the surface of the pubic symphysis, some numerical representation of its shape needs to be provided. Three-dimensional (3D) laser scanners generate a dense point cloud or a polygonal mesh that captures the geometry of a physical object with hundreds or thousands of coordinate (x, y, z) measurements. The 3D shape is represented as numerous small adjacent triangles, called faces. The data files produced by the 3D scanners store the coordinates of all the vertices and the information on how they are connected to form the triangles. The software we developed uses ASCII and binary PLY (<http://paulbourke.net/dataformats/ply/>) files.

3D laser scan samples

To develop and test the method proposed in this article, 3D laser scans of the pubic symphysis from 56 cases were used. This sample was sourced from a larger collection of bracket scans (scanned at three consecutive angles) taken with the NextEngine 3D Desktop Scanner (Model 2020i). Each pubic bone was placed approximately 5 to 9 inches away from the scanner box and oriented such that the face of the symphysis was positioned at 90° or perpendicular to the base of AutoDrive scanning stand, onto which the ramus was secured with modeling clay (see Fig. 1). Individual scans were produced in approximately 3.5 min, using the following high definition (HD) scanning settings: neutral image capture, 16 divisions, 40 K points per inch² with a triangle size of 0.0050. These scans were manipulated using the accompanying software, Scan Studio HD version 1.3.2, such that multiple scans were autoaligned to form a single mesh, the face of the pubic symphysis mesh was isolated by visual inspection and the surrounding areas of the bone deleted, leaving only the trimmed surface mesh for the region of interest. Postprocessing steps, like hole filling, smoothing and simplification, were not



Fig. 1. Position and orientation of the pubic bone during the scanning of the pubic symphysis.

performed in order to preserve the raw qualities of the scan for ease of future replication. These data collection protocols were repeated by two (senior and junior) observers at different times and locations for a random set of cases to assess the degree of variation in the meshes produced and the ages estimated using our algorithm (see the Results section). From the final aligned and trimmed meshes, we extracted the 3D coordinate data (x , y , z values) that characterize the surface of the pubic symphysis for each sample and subjected these values to our method's computational procedure.

Our primary sample includes a single scan for each of 44 modern American individuals, opportunistically selected from the more recent cases available in the W.M. Bass Skeletal Collection, curated by the Forensic Anthropology Center at the University of Tennessee in Knoxville (<http://fac.utk.edu/>). As some studies have shown that source population and sex may affect the morphology of the pubic symphysis as well as its age-progressive changes (Gilbert and McKern, 1973; Katz and Suchey, 1986, 1989; Hoppa, 2000; Berg, 2008; Kimmerle et al., 2008), the individuals used here represent self-identified white males of known age-at-death and birth year. Candidates were not screened for pathological conditions, but the preservation and completeness of the os coxae was of concern to ensure that alternative methods could be applied in order to produce comparative age data. Efforts were made to include as many young and very old individuals as possible given the constraints imposed by the Bass Collection's composition of largely older aged donors. This schema allowed us to ensure that a large proportion of the sample could also

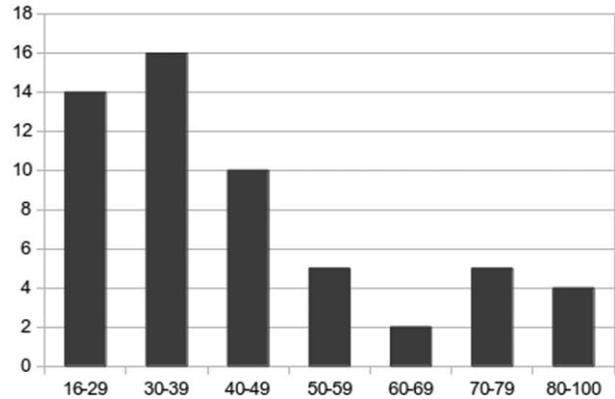


Fig. 2. The age distribution for the 56 scans.

be aged using conventional estimation procedures without needing to assign the skeletons to a terminal, open-ended interval of 50+ years (following the example of Milner and Boldsen, 2012) and to approximate the age distribution of the cases used in the development of the Suchey-Brooks method (see Brooks and Suchey, 1990 for their sample histogram). Figure 2 shows the age distribution of our sample. While we have considered additional cases ($n > 150$), we report here on this reduced sample of 44 so that we refer to the same cases analyzed by Slice and Algee-Hewitt (2015), whose work sets the precedent for our computationally driven, laser scan approach. Examples of four of the scans of different ages-at-death included in this analysis are given in Figure 3. The structure of the scans, vertices, and connected face edges, is illustrated in Figure 4. In addition to the 44 Bass Collection scans, 12 casts representing the Suchey-Brooks phases were scanned and added to the dataset. Each of the six phases is represented by two cast scans, with an associated age. These casts serve as our control samples: as they are publicly accessible, their scans can be reproduced and our analysis externally evaluated.

Standardizing the orientation

In order to apply the same algorithm to all of the scans, we need to first standardize their position, size, and orientation. Standardization of the position for each scan is done by first translating the set of vertices so that the middle point of the x , y , and z coordinates matches the center of the coordinate system. This is achieved by simply subtracting the middle values from the x , y , and z coordinates of the entire set of vertices from the coordinates of each vertex. Once the scan is positioned at the origin, a principal component analysis (PCA) is used to rotate it so that the x , y , and z coordinates measure the largest variances in each dimension. PCA is an orthogonal, linear transformation that rotates the bone to a new coordinate system. As a result of the rotation the x -axis defines the dimensions with the largest variance, the y -axis defines the dimension with the second largest variance, and the z -axis defines the dimension with the smallest variance. This way the x - y plane approximates the articular surface of the bone, and we model variation in z direction. Figure 5 shows, in two different orientations, an example of the pubic symphysis after the rotation. PCA incorporates no information on anatomical orientation.

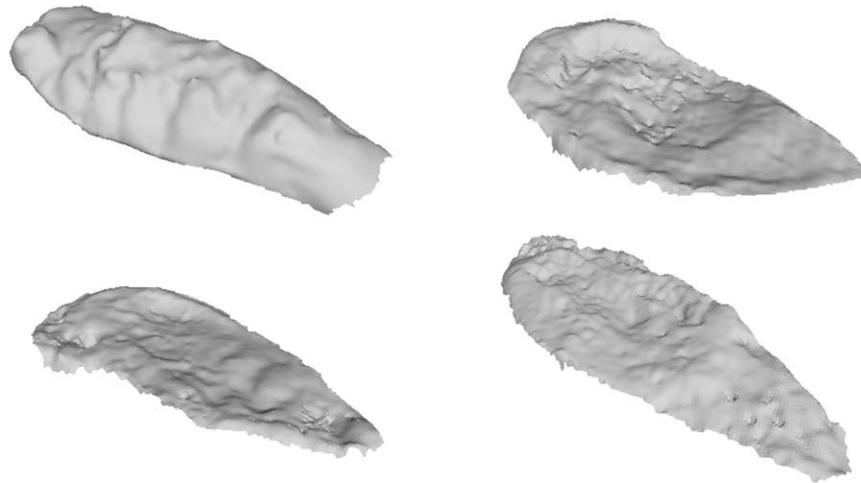


Fig. 3. Four examples of the 3D scans used in the model. The exact ages-at-death starting from the top left scan and moving in a clockwise direction are 19, 39, 59, and 96 years.

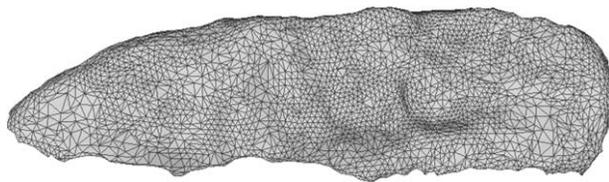


Fig. 4. An example of a 3D scan showing the adjacent triangles representing the curves on the surface.

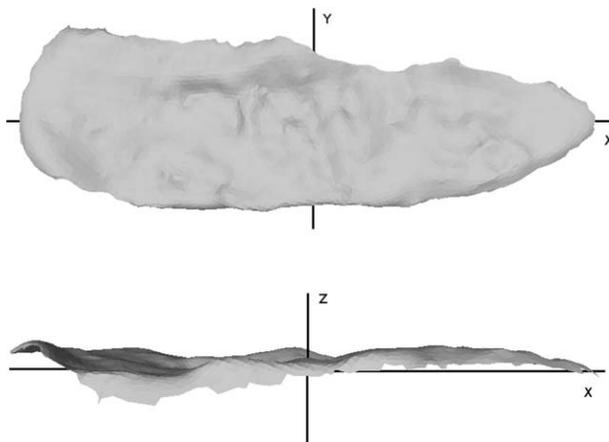


Fig. 5. The position of the 3D scan after the translation to the center of the coordinate system and PCA rotation. The figure shows the scan from two different angles with respect to the coordinate axes.

To ensure that scans will be of uniform size, we multiply all of the vertex coordinates of each scan by a scaling factor. The scaling factor is estimated for each scan using the current size of the scan and the final desired size. In order to preserve the aspect ratio of the scans only the size in one of the dimensions (in our case the length) is used in the calculation of the scaling factor. The Appendix offers details on implementing the affine transformations.

The PCA positioning aligns the length, width, and depth of the scans but we also orient the anatomical parts of the pubic symphysis. This is done manually using a visualization tool included in the software that allows the users to see the scans and flip them around the axes. The individuals included in our study are represented by a scan of their left or right pubic symphysis but not both. Flipping around the axis allows for all of the scans, both left and right, to be oriented uniformly. For our method, we positioned all scans in such a way that the superior apex is to the left (negative x coordinates), the ventral margin is on top (positive y coordinates), and the dorsal margin is at the bottom (negative y coordinates). The anatomical orientation of the scans does not affect the computational method as long as the same sample points are used.

The thin plate splines (TPS)

In this article, we adopt the sophisticated surface characterization technique of thin plate splines (TPS). The TPS algorithm models the bending of an infinitely thin, flat metal plate (Bookstein, 1989). In the case of the 3D laser scans, the hypothetical metal plate will be bent to match the surface of the pubic symphysis. The method computes the minimum energy that is required for this bending. Our expectation is that the changes occurring on the symphyseal face, transforming from a high billowed surface to a smoother, flattened one, will be captured by the method as the required minimum bending energy decreases. This novel approach to the quantification of the pubic symphyseal surface allows us to address acknowledged limitations of variance-based methods, most notably that the spatial structure of the data is not taken into account and similar results can be returned for different morphological expressions simply because they have the same variance (Slice and Alge-Hewitt, 2015). When applied to the different graphs in Figure 6, a variance-base method will return the same value for both because they both have the same variance, but the minimum bending energy (BE) required for the plot on the left will be higher.

The first step of the TPS algorithm requires the selection of two sets of control points. One set of points lies on the plane that is bent and the other on the bone

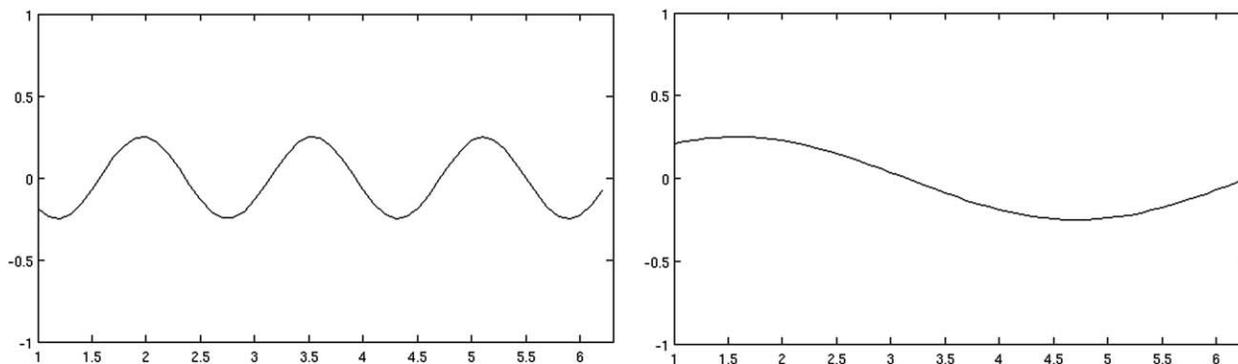


Fig. 6. Examples of surfaces that have the same variance but a higher minimum bending energy (left) and lower minimum bending energy (right).

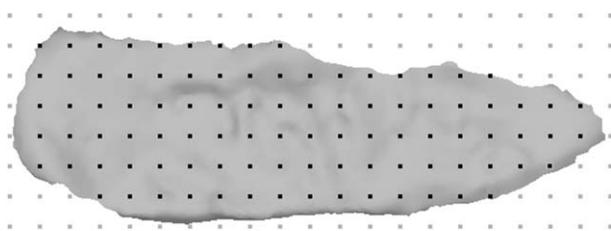


Fig. 7. The uniform square mesh used by the method plotted against the corresponding scan. The grey vertices are considered to be outside of the surface and are not included in the calculations.

surface. The z values of the control points on the plane need to be mapped to the z values on the points of the bone. The exact mapping of the control points involves solving a linear system $A\vec{x} = \vec{b}$. The solution of the system \vec{x} is used to interpolate the rest of the points and to estimate the minimum bending energy.

Selecting control points. We begin by creating a grid of square cells on a plane that extends from the minimum to the maximum values in the x and y dimensions of the scans. A number of points in one of the x or y dimensions is selected (x was arbitrarily chosen). The points divide the length of the scan into equal intervals. Then enough points in the y direction are selected that are the same distance apart from each other, to cover the width of the scan. All the points are assigned a z value of 0. As a result of the PCA rotation, this plane passes through the scan.

The shape of the bones is, however, not rectangular. When the initial grid overlaps the scan, some of the grid points will be outside of the scan area. The TPS algorithm matches two sets of control points exactly and, since some of the points on the grid do not have a corresponding point on the scan, the extraneous points need to be identified and removed. To implement this, all the vertices are projected onto the x - y plane simply by taking their z -coordinates to be 0, i.e., they are treated as 2D objects. This means that all the initial mesh points and all the triangle projections lie on the same plane and each mesh point can be inside only one triangle since all occlusions on the scans were removed during the cleaning process. For each mesh point the algorithm

finds the face for which it lies inside the projection. If a mesh point is not inside any triangle this means that it lies outside of the scan and is removed from the set of control points. Figure 7 illustrates the process of selecting the control points on the plane. It shows the uniform square mesh with the outside points colored in gray and the final set of control points colored in black. At this point each mesh point is associated with one face. The equation for the 3D plane containing the three face vertices (not their 2D projections) is constructed and the projection of the mesh point onto that plane is calculated. The set of projected mesh points becomes the second set of control points. The points on the two control sets have the same x and y values but different heights (z values). The heights will be mapped by the thin plate splines algorithm.

TPS algorithm. Let the selected vertices on the surface scan have the coordinates b_x^i , b_y^i and b_z^i , where i denotes the index of the vertex. If there are k selected vertices, then $1 \leq i \leq k$. Let the coordinates of the selected points on the plane be denoted as p_x^i , p_y^i , and p_z^i . Let C be a $k \times k$ matrix such that

$$C = \begin{pmatrix} U(r_{11}) & U(r_{12}) & U(r_{13}) & \cdots & U(r_{1k}) \\ U(r_{21}) & U(r_{22}) & U(r_{23}) & \cdots & U(r_{2k}) \\ U(r_{31}) & U(r_{32}) & U(r_{33}) & \cdots & U(r_{3k}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U(r_{k1}) & U(r_{k2}) & U(r_{k3}) & \cdots & U(r_{kk}) \end{pmatrix},$$

where r_{ij} , for $1 \leq i \leq k$ and $1 \leq j \leq k$, is the distance between point i and point j in the plane. The function U is called a basis function, and is defined, for planar data, as:

$$U(r) = 0, \text{ for } r = 0 \\ U(r) = r^2 \log(r), \text{ for } r \neq 0$$

The distance between any point and itself is equal to 0, therefore $U(r_{ij}) = 0$ when $i = j$ so that, in fact, all the diagonal entries of matrix C are 0.

Let P be a $k \times 3$ matrix such that

$$P = \begin{pmatrix} 1 & p_x^1 & p_y^1 \\ 1 & p_x^2 & p_y^2 \\ \vdots & \vdots & \vdots \\ 1 & p_x^k & p_y^k \end{pmatrix}.$$

And let M be a $(k+3) \times (k+3)$ matrix such that

$$M = \begin{pmatrix} C & P \\ P^T & 0 \end{pmatrix}.$$

The right hand side, \vec{b} , of the linear system is of size $k+3$ and is obtained from the z coordinates of the selected points on the bone, i.e.,

$$\vec{b} = \begin{pmatrix} b_z^1 \\ b_z^2 \\ \vdots \\ b_z^k \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Now that the matrix M and the right hand side \vec{b} of the linear system $M\vec{x} = \vec{b}$, are constructed, it can be solved for \vec{x} .

Finding the inverse of a matrix is time consuming and computationally expensive. To improve the efficiency of the method, instead of finding the inverse of M , an LU factorization with permutations is used. It factors the matrix M into a lower triangular matrix L and an upper triangular matrix U by switching the rows of M and \vec{b} in a consistent way. Having the L and the U matrices, we can avoid computing the inverse of the matrix by solving forward and backward systems (Kinkaid and Cheney, 2002).

Now that \vec{x} is known, its components can be used to interpolate the height of every point on the plane. The vector \vec{x} is of size $k+3$. Let \vec{w} denote the first k elements of \vec{x} and \vec{a} denote the last three elements. Then,

$$p_z^l = a_1 + a_2 p_x^l + a_3 p_y^l + \sum_{i=1}^k w_i U \|p^i - p^l\|_2,$$

where p^l is any point on the plane. After the transformation the value of p^l will approximate the value of a point on the bone if it is not one of the control points. If p^l is one of the control points, its value will match exactly the z value of the corresponding control point on the bone. The construction of the actual approximation of the surface is not, however, needed for age estimation. The method uses the minimum bending energy and only the solution \vec{x} of the linear system is required to calculate it.

Minimum bending energy

The value of the minimum bending energy, E_{\min} , is given by

$$E_{\min} = \vec{w} C \vec{w}^T$$

As we have pointed out earlier, the TPS algorithm models the bending of a metal plate to match another surface. Some force needs to be applied to the metal plate to transform it. The minimum bending energy, as the term suggests, is the least amount of energy required to transform the infinitely thin flat plate into the shape of the surface scan. It minimizes the integral

$$\iint \left(\left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 \right) dx \, dy$$

(Bookstein, 1991). This minimum bending energy is a measure of the complexity of the surface scan. If an association can be made between age-at-death and the BE, this relationship could be used to estimate the age-at-death of a given individual. For the analysis and our regression model we used logarithmic functions, $\log(\text{age})$ and $\log(\text{bendingenergy})$, to enhance the linearity of the relationship.

Software development and procedure

The execution of the method includes a set procedure. Step 1: the PLY file data provided by the 3D scan is read. Step 2: the bone is subjected to a translation followed by PCA to standardize positioning, scaling to standardize the size and manually orienting the anatomical parts. Step 3: the control points (a set of points on a flat plane and on the bone, which are to be matched exactly) are selected. Step 4: the TPS algorithm is implemented, and the minimum bending energy (the energy necessary to deform the planar landmarks to exactly match those on the scan surface) is computed. The software implementing the algorithm, forage, is available for free download as a standalone Java application from <http://morphlab.sc.fsu.edu>.

RESULTS

Our TPS method was tested using pubic bone scans from a sample of 44 modern American male skeletons for whom age-at-death was known and 12 scans of casts representing the six phases of the Suchey-Brooks method. After successful execution of our algorithm, we obtained 56 minimum bending energy values. The linear relationship between the age-at-death and the minimum bending energy was studied.

Correlation and linear regression analysis on the relationship between the $\log(\text{age})$ and the $\log(\text{bendingenergy})$ were performed. The coefficients for the two analyses are -0.4765 and -0.2730 , respectively. A p -value of 0.0002 for both confirms the statistical significance of our findings at the usual p -value less than 0.05 level. The equation for the regression line is $y = -0.2730x + 3.9537$. In our case y is $\log(\text{age})$ and x is $\log(\text{bendingenergy})$. The age approximation is 10^y . The root mean square error (RMSE) for exact and estimated ages is estimated to be 18.6183 years using the formula

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{estimated}_i - \text{exact}_i)^2},$$

where n is the number of bone scans that are used in the model. The plot in Figure 8 shows the regression line, the 12 cast scans of the Suchey-Brooks phases shown as filled circles and the rest of the 44 scans

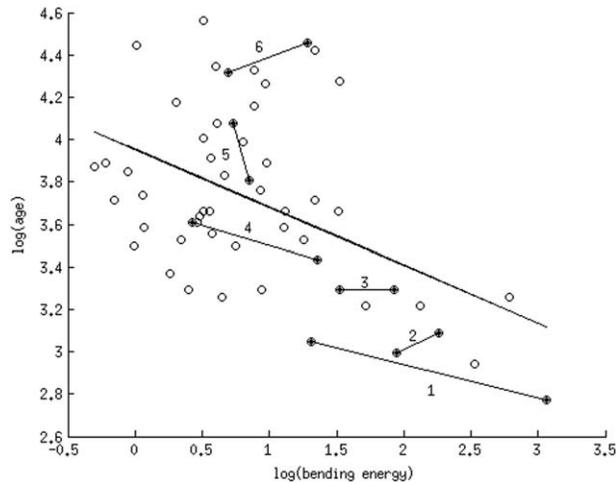


Fig. 8. Results and regression line using the surface scans, p -value = 0.0002, $R^2 = 0.2270$, RMSE = 18.6183 years. The Suchey-Brooks casts are shown as filled circles. The two casts in each phase are connected with a line and the corresponding phase numbers appear by the connecting lines.

shown as empty circles. The two cast scans representing each phase are connected with straight lines and the number of the corresponding phase is listed by the line.

The findings reported above were generated using a mesh with 31 points in the x direction. This mesh size was arbitrarily chosen for testing purposes. The results of using meshes of different size, ranging from 11 to 101 points in x direction, were compared, and Table 1 shows the p -value, the coefficient of determination (R^2) and the RMSE in years using different number of control points. The p -value fluctuates around 0.002 as the number of control points increases and the RMSE fluctuates slightly around 19 years. The coefficient of determination shows that for the grid sizes producing the smallest RMSEs and p -values, about 20 to 25% of the age variation is explained by the minimum bending energy.

Tables 2 and 3 show in detail the results for the Suchey-Brooks casts and the Bass collection bones/scans, respectively. The exact ages-at-death for the casts were provided by Judy Suchey and the phase age-ranges reported in Table 2 were taken from the original article (Brooks and Suchey, 1990). Table 3 presents true ages relative to those estimated from the laser scans using our algorithm as well as the scores assigned to the corresponding bones by a skeletal biologist experienced in using the Suchey-Brooks method and very familiar with the morphological variation characteristics of the Bass Collection.

To check the consistency of the method and to test its behavior in practice, we applied a leave-one-out cross validation procedure, such that a) each of our specimens was removed from the data set, b) the model was built using the remaining data, and c) the age of the excluded specimen was predicted with the new model to which it did not contribute. The RMSE for the leave-one-out cross validation analysis using a mesh with 31 points in the x direction is 18.9619.

As discussed in the Materials and Methods section a random sample of 12 bones were scanned and cleaned three times independently for a total of 36 scans. The bending energy was estimated for each of the 36 scans and the average of the three values for each bone was used in the results reported so far in this article. Three

TABLE 1. List of regression P values, R^2 , and root mean square errors (RMSE) in years using meshes of different sizes (11 through 101 points in x -direction) for the surface scans

Mesh size	P	R^2	RMSE
11	0.0001	0.2407	18.2160
21	0.0001	0.2597	18.3498
31	0.0002	0.2270	18.6183
41	0.0002	0.2239	18.6140
51	0.0005	0.2037	18.7417
61	0.0070	0.1271	19.0441
71	0.0061	0.1311	19.1512
81	0.0052	0.1357	19.0160
91	0.0037	0.1459	18.8807
101	0.0046	0.1390	18.9539

TABLE 2. Results for the 12 Suchey-Brooks casts: the table shows the corresponding Suchey-Brooks phase, the exact age-at-death, TPS age-estimate, and Suchey-Brooks phase age-range

Suchey-Brooks phase	Exact age	TPS estimated age	Phase range
I-I	16	23	15–23
I-II	21	36	
II-I	22	28	19–34
II-II	20	31	
III-I	27	31	21–46
III-II	27	34	
IV-I	31	36	23–57
IV-II	37	46	
V-I	59	43	27–66
V-II	45	41	
VI-I	75	43	34–86
VI-II	86	37	

additional regression models were built using the three sets of bone scans. The age estimates for each bone using the three sets of scans differed by, at most, 3 years. In addition the differences between the p -values and the RMSEs were trivial. Analysis of variance results suggest that the bending energies of individual bones are significantly different even when accounting for the difference in scanning context and practitioner-experience. The results are statistically significant at the 1% level.

DISCUSSION

The results of the thin-plate splines algorithm presented in this article are comparable to the well-established and preferred Suchey-Brooks system. Our method is being actively expanded and has the potential to become more accurate but, even in its current state, it reduces the high concern over uncontrolled observer error and subjectivity. This methodological improvement gives us the ability to produce high quality estimates that will contribute broadly to osteological laboratory analysis and paleodemographic inference but also to align skeletal methods for age-estimation with the probabilistic approaches that are of growing significance in other areas of forensic case analysis, such as predictive phenotypic models and profile matching for individual identification in genetics. With computational methods like ours we can now more effectively respond to call of improved standards in the forensic sciences.

Our results show that the deep billowing that characterizes early stages of aging is best captured by the thin

TABLE 3. Results of exact age, TPS age estimate, assigned Suchey-Brooks phase, and the corresponding phase age interval for each of the 44 W.M. Bass Skeletal Collection scans/bones used in the model

Exact age	TPS estimated age	Suchey-Brooks phase	Phase range
19	26	I (late)	15–23
25	29	II (late)	19–34
25	33	I (late)	15–23
26	44	IV (early)	23–57
26	24	II (late)	19–34
27	47	IV (late)	23–57
27	40	III (early)	21–46
29	48	IV (late)	23–57
33	42	IV (early)	23–57
33	52	IV (late)	23–57
34	47	IV (early)	23–57
34	37	IV (late)	23–57
35	45	IV (late)	23–57
36	39	IV (late)	23–57
36	51	IV (late)	23–57
37	46	IV (early)	23–57
37	46	IV (late)	23–57
38	46	IV (late)	23–57
39	45	V (late)	27–66
39	38	IV (early)	23–57
39	45	IV (late)	23–57
39	35	V (early)	27–66
41	36	V (late)	27–66
41	54	V (late)	27–66
42	51	V (early)	27–66
43	40	V (early)	27–66
46	43	V (early)	27–66
47	53	V (late)	27–66
48	57	V (early)	27–66
49	40	V (early)	27–66
49	55	V (early)	27–66
50	45	VI (early)	34–86
54	42	IV (late)	23–57
55	45	VI (early)	34–86
59	44	V (early)	27–66
64	41	VI (early)	34–86
65	48	VI (early)	34–86
71	40	VI (late)	34–86
72	34	VI (late)	34–86
76	41	V (late)	27–66
77	44	VI (early)	34–86
83	36	V (late)	27–66
85	52	V (early)	27–66
96	45	VI (late)	34–86

plate splines algorithm. For later age stages, as the pubic symphysis face flattens and develops more complex irregular characteristics such as porousness, lipping, and pitting (Brooks and Suchey, 1990), its shape is more difficult to evaluate. Furthermore, the rim, breakdown of the margins and the occasional gap that exists on the top ventral half of the scans (Milner and Boldsen, 2013) result in high bending energy values for older people so the ages are underestimated. Accordingly, looking at the results reported in Tables 2 and 3, we can see that the age estimates generated by our algorithmic approach are much better for younger individuals. Figure 8 shows that phases 1, 2, 3, and 4 of the Suchey-Brooks casts generally fall along the regression line in the expected order. However, phases 5 and 6 overlap phase 4 and cannot be distinguished using the bending energy values. Overall our age estimates fall within the corresponding Suchey-Brooks phase, but it should be noted that the ages for some of the

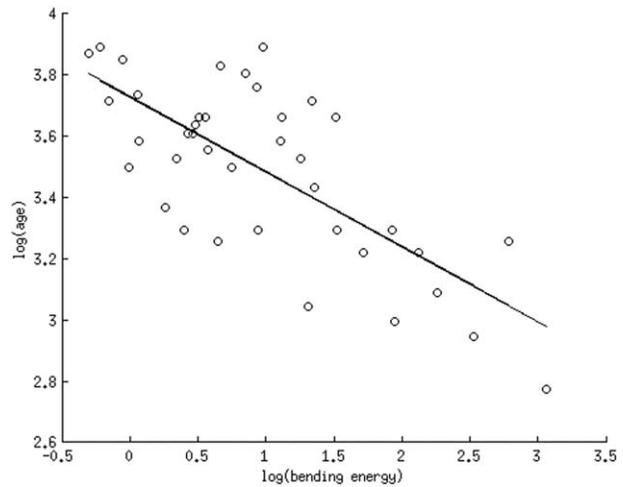


Fig. 9. Results and regression line using only scans for people who are younger than 50 years. The number of data points is 40 and the RMSE is 6.39 years.

younger individuals are overestimated due to the regression line being strongly influenced by the older individuals. Even though the age estimates for the older males are highly underestimated, the large age intervals associated with later Suchey-Brooks phases allow for our results to still be significant. To confirm our observation that the method performs better when applied to younger individuals we built a model using only the 40 males with exact age-at-death less than 50 years. Figure 9 shows the data and regression line for the new model. The RMSE is only 6 years.

FUTURE DIRECTIONS

Building upon the advances made by Milner and Boldsen (2013) in their component-driven approach to indicator scoring, we are actively extending this research to include analysis of the different pubic symphysis parts. Preliminary tests on the dorsal and the ventral halves separately promise good results. This kind of partial analysis will be useful in situations when the entire face of the pubic symphysis is not available for study due to loss of bony macrostructure not associated with aging, like fragmentation from trauma or postmortem damage, and necessary exclusion on account of pathological conditions. Further, we foresee success in the method's application to other age indicators such as the sternal rib ends, and the medial surface of the clavicle, given their topographic and developmental/degenerative similarities. The model presented in this article was developed using only self-identified white American males. Its performance needs to be tested using mixed-sex and multipopulational data. Preliminary work on an unpublished sample of South-West Hispanics yielded positive results. The Hispanic data is unidentified so only Suchey-Brooks scores are available as age estimates but 8 out of the 11 males had an age estimate that accurately mapped to the corresponding phase age-range.

To ensure accessibility to this approach, we developed software that enables the straightforward execution of our TPS method by skeletal biologists and medicolegal practitioners. The software is free for download (<http://morphlab.sc.fsu.edu>).

CONCLUSION

Our tests of the thin plate spline (TPS) algorithm method are encouraging for resolving the ongoing concerns over method error reduction and standardization. We have demonstrated that it is possible to generate reliable and accurate estimates of age-at-death in a way that is invariant to the level of expertise and experience of the practitioner. This article's results show that there exists a significant linear relationship between the value of the log(minimum bending energy) generated by the TPS algorithm and the log(age-at-death) of the individual. When the TPS method was applied to the entire scan surface, we identified a significant correlation, which resulted in a root mean square error of approximately 19 years and a p -value of 0.0002 when using a grid with 31 points in x direction. These findings demonstrate that given the scan of a new pubic bone, our method can be used with success to calculate the minimum bending energy associated with it: after which, the equation for the line describing the relationship between the two variables can be used to estimate the age. Overall, the method was consistent in estimating high bending energy for younger cases and lower bending energy and for older cases.

It is important to note that we have confirmed that our TPS approach can yield estimates comparable to the established morphoscopic age-at-death techniques, including Suchey-Brooks method. Our method offers, however, the added advantage of a fully integrated, objective and quantitative, framework of analysis that has potential for future refinement and increased estimation accuracy. Further, the design of our morphometric approach offers many practical improvements. The full-workflow (from laser scan acquisition to age estimation) allows for the ongoing incorporation of reference samples that represent different geographic, temporal, and socially defined populations that are of archaeological and forensic significance. As scanner data is stored in a file and can be used to reconstruct the skeletal remains in virtual space, 3D models can be archived for future study and downstream statistical analyses can be performed in the absence of the physical materials.

ACKNOWLEDGMENTS

The authors thank the members of the Forensic Anthropology Center and the Department of Anthropology at the University of Tennessee, who facilitated access to the modern skeletal collection and laboratory equipment. The authors are most grateful to Jieun Kim for providing them with the Suchey-Brooks scores and sharing many thoughtful suggestions. They also thank Dr. Bruce Anderson for providing access to PCOME cases and Dr. Cris Hughes for assisting with the laser scanning of these individuals.

APPENDIX

AFFINE TRANSFORMATIONS APPLIED TO THE SCANS

Translation

Translation is essentially an operation that moves a given set of points a constant distance in a given direction. In order to translate the scan to the center of the coordinate system, it is necessary to calculate the current position of the geometric center of the scan. This is done by finding the middle point between the smallest and the largest values in each of the three dimensions.

This point is used as the current center of the scan. Then, the equivalent of the following matrix-vector multiplication is carried out:

$$\begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \\ 1 \end{pmatrix} = \begin{pmatrix} p_x - c_x \\ p_y - c_y \\ p_z - c_z \\ 1 \end{pmatrix},$$

where c_x , c_y , and c_z are the x , y , and z coordinates of the current position of the center of the scan and p_x , p_y , and p_z are the coordinates of each of the n vertices of the scan. The vertices are now centered with their middle point at (0,0,0).

Principal component analysis

Once the bone is positioned at the origin, PCA is used to rotate it so that the x , y , and z axes define the dimensions with the largest variance (x), the second largest variance (y), and the smallest variance (z). PCA is an orthogonal, linear transformation that rotates the bone to a new coordinate system. In the current application, the PCA algorithm uses a centered data matrix, A , that is an $n \times 3$ matrix where n is the number of vertices and the three columns correspond to the x , y , and z coordinates of each vertex after the centering.

First, the 3×3 matrix defined by $A^T A$ is calculated. Then, this matrix is factorized using a singular value decomposition (SVD) (Kinkaid and Cheney, 2002). The $A^T A$ matrix is factored into a 3×3 diagonal matrix S that has the singular values as its diagonal entries and two orthogonal 3×3 matrices U and V , such that $A^T A = USV^T$. Since $A^T A$ is symmetric and positive definite, $U = V$. The $n \times 3$ matrix AV gives the coordinates for the vertices after the PCA rotation.

Scaling

To make the scans of uniform size, it is necessary to multiply all of the vertex coordinates of all of the scans by some scaling factor that is different for the different scans. Not all three dimensions can be of equal size because this will distort the aspect ratio of the scans. The method scales the length x of the scans uniformly for all of them and changes the width y and depth z accordingly. Finding the scaling factor requires finding the current length (the distance between the smallest and the largest values in x direction after PCA) of the scan (Δx_0) and selecting a length for the scan after the scaling (Δx_1). For our method we scaled all scans to be of unit length. The scaling factor S is the ratio of the two numbers ($\frac{\Delta x_1}{\Delta x_0}$). Then, the equivalent of the following matrix-vector multiplication is carried out:

$$\begin{pmatrix} S & 0 & 0 & 0 \\ 0 & S & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \\ 1 \end{pmatrix} = \begin{pmatrix} S \times P_x \\ S \times P_y \\ S \times P_z \\ 1 \end{pmatrix},$$

where p_x , p_y , and p_z are the coordinates of each vertex of the scan.

LITERATURE CITED

- Algee-Hewitt, BFB. 2013. Age estimation in modern forensic anthropology. In: Tersigni-Tarrant MT, Shirley NR, editors. *Forensic anthropology: an introduction*. Boca Raton, FL: CRC Press. p 181–230.
- Baccino E, Ubelaker DH, Hayek LA, Zerilli A. 1999. Evaluation of seven methods of estimating age-at-death from mature human skeletal remains. *J Forensic Sci* 44:931–936.
- Berg GE. 2008. Pubic bone age estimation in adult women. *J Forensic Sci* 53:569–577.
- Bookstein F. 1989. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell* 11:567–585.
- Bookstein F. 1991. *Morphometric tools for landmark data*. New York, NY: Cambridge University Press.
- Brooks ST. 1955. Skeletal age at death: the reliability of cranial and pubic age indicators. *Am J Phys Anthropol* 13:567–589.
- Brooks S, Suchey JM. 1990. Skeletal age determination based on the OS pubis: a comparison of the Acsadi-Nemeskeri and Suchey-Brooks methods. *Hum Evol* 5:227–238.
- Buikstra JE, Ubelaker DH. 1994. *Standards for data collection from human skeletal remains: proceedings of a seminar at the Field Museum of Natural History*. Fayetteville, AR: Arkansas Archeological Survey.
- Christensen AM, Crowder CM. 2009. Evidentiary standards for forensic anthropology. *J Forensic Sci* 54:1211–1216.
- Garvin HM, Passalacqua NV. 2012. Current practices by forensic anthropologists in adult skeletal age estimation. *J Forensic Sci* 57:427–433.
- Gilbert BM, McKern TW. 1973. A method for aging the female OS pubis. *Am J Phys Anthropol* 38:31–38.
- Hartnett KM. 2010. Analysis of age-at-death estimation using data from a new, modern autopsy sample - part I: pubic bone. *J Forensic Sci* 55:1145–1151.
- Holden C. 2009. Forensic science needs a major overhaul, panel says. *Science* 323:1155–2009.
- Hoppa RD. 2000. Population variation in osteological aging criteria: an example from the pubic symphysis. *Am J Phys Anthropol* 111:185–191.
- Hoppa RD, Vaupel JW. 2002. *Paleodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press.
- Katz D, Suchey JM. 1986. Age determination of the male OS pubis. *Am J Phys Anthropol* 69:427–435.
- Katz D, Suchey JM. 1989. Race differences in pubic symphyseal aging patterns in the male. *Am J Phys Anthropol* 80:167–172.
- Kimmerle EH, Jantz RL, Konigsberg LW, Baraybar JP. 2008. Analysis of age-at-death estimation through the use of pubic symphyseal data. *J Forensic Sci* 53:558–568.
- 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.
- Kinkaid D, Cheney W. 2002. *Numerical analysis: mathematics of scientific computing*. Pacific Grove, CA: Brooks/Cole.
- 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.
- McKern TW, Stewart TD. 1957. *Skeletal age changes in young American males analysed from the standpoint of age identification*. Technical report EP-45. Natick, MA: Quartermaster Research and Development Command.
- Meindl RS, Lovejoy CO, Mensforth RP, Walker RA. 1985. Method of age determination using the OS pubis, with a review and tests of other current methods of pubic symphyseal aging. *Am J Phys Anthropol* 68:29–45.
- Milner GR, Boldsen JL. 2012. Transition analysis: a validation study with known-age modern American skeletons. *Am J Phys Anthropol* 148:98–110.
- Milner GR, Boldsen JL. 2013. *Transition analysis age estimation: skeletal scoring manual*. Fordisc Version 1.00. Available at: <http://anth.la.psu.edu/research/bioarch/docs/TAMannual2013-June.pdf>. Last accessed June 17, 2015.
- Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schütz HW, Kaatsch HJ, Borrman HI. 2000. Age estimation: the state of the art in relation to the specific demands of forensic practice. *Int J Legal Med* 113:129–136.
- Scientific Working Group for Forensic Anthropology (SWGANTH). Age Estimation 2013. <http://swganth.org/>. Last accessed June 17, 2015.
- 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*. Apr 30 (early view online)
- Suchey JM, Katz D. 1998. Applications of pubic age determination in a forensic setting. In: Reichs KJ, editor. *Forensic osteology: advances in the identification of human remains*, 2nd ed. Springfield, IL: Charles C. Thomas. p 204–236.
- Suchey JM, Wiseley DV, Katz D. 1986. Evaluation of the Todd and McKern-Stewart methods for aging the male OS pubis. In: Reichs KJ, editor. *Forensic osteology – advances in the identification of human remains*. Springfield, IL: Charles C. Thomas. p 33–67.
- Todd TW. 1920. Age changes in the pubic bone. *Am J Phys Anthropol* 3:285–327.

ANTHROPOLOGY

Detelina K. Stoyanova,^{1,†} Ph.D.; Bridget F. B. Algee-Hewitt,^{1,2,3,†} Ph.D.; Jieun Kim,^{1,3} Ph.D.; and Dennis E. Slice,^{1,4} Ph.D.

A Computational Framework for Age-at-Death Estimation from the Skeleton: Surface and Outline Analysis of 3D Laser Scans of the Adult Pubic Symphysis*

ABSTRACT: In forensic anthropology, age-at-death estimation typically requires the macroscopic assessment of the skeletal indicator and its association with a phase or score. High subjectivity and error are the recognized disadvantages of this approach, creating a need for alternative tools that enable the objective and mathematically robust assessment of true chronological age. We describe, here, three fully computational, quantitative shape analysis methods and a combinatory approach that make use of three-dimensional laser scans of the pubic symphysis. We report a novel age-related shape measure, focusing on the changes observed in the ventral margin curvature, and refine two former methods, whose measures capture the flatness of the symphyseal surface. We show how we can decrease age-estimation error and improve prior results by combining these outline and surface measures in two multivariate regression models. The presented models produce objective age-estimates that are comparable to current practices with root-mean-square-errors between 13.7 and 16.5 years.

KEYWORDS: forensic science, age-at-death estimation, pubic symphysis, age indicator morphology, biological profile, skeletal casework analysis, morphometrics, 3D laser scans, outline analysis, multivariate regression, bone surface mapping

Identifying the most appropriate method for estimating age-at-death from adult skeletal remains has inspired sustained interest in methodological innovation among forensic anthropologists, in particular over the last three decades (1). This investment in age-at-death estimation reflects its importance to skeletal analysis as, with other demographic parameters, like sex, ancestry, and stature, it represents an essential component of the biological profile, providing key information for individual identification in medico-legal casework (1). Given the established history of research and publication on the pubic symphysis by, for example, Todd (2), McKern and Stewart (3), Suchey and Brooks (4–7), and, most recently, Milner and Boldsen (8), it is not surprising that, despite the existence of other age-informative regions in the skeleton (e.g., auricular surface, sternal rib ends, cranial sutures, and the medial surface of the clavicle), it remains the most reliable (9), most frequently used (10), and the most widely preferred (11) of the indicators considered in age-at-death estimation. While there have been recent efforts to refine how

the pubic symphysis is visually assessed and to define new statistical approaches to its analysis (12–16), standard practice continues to require the macroscopic comparison of the bone surface morphology of the pubic symphysis to a set of population-specific criteria that represent a series of pre-defined phases, such that the case-specific age-at-death can be estimated from an age-range previously associated with the assigned phase (2,3,5). While this approach is attractive for its simplicity, its fundamental processes of feature-scoring and bone-to-phase-matching are methodologically subjective. It is also questionably appropriate for remains whose full review of the bone surface is obscured by factors such as fragmentation, pathologies, trauma, or taphonomic defects. Its efficacy rests, therefore, upon the training and experience of the practitioner, the degree of preservation, and the morphological typicality of that specimen relative to the method-specific standards, which may include text-based descriptions, photographs, and comparative casts (17). While the problems of method and observer error inherent within this kind of skeletal age-estimation protocol are well-documented across the field (18–20) and important work has been done on advancing additional, notably component-based, scoring methods (14), forensic anthropology currently lacks an alternative to this core “macro-morphoscopic” procedure. There is, therefore, a need to broaden our thinking on how we can differently and more productively assess the structure of the pubic symphysis, and, ultimately, harness this information to infer true chronological age with reduced error. Here, we advocate for developing new *objective, quantitative* methodologies that not only resolve these pre-existing limitations but are also satisfying to the evolving

¹Department of Scientific Computing, Florida State University, Tallahassee, FL 32306.

²Department of Biology, Stanford University, Stanford, CA 94305.

³Department of Anthropology, Florida State University, Tallahassee, FL 32306.

⁴Department of Anthropology, University of Vienna, Vienna 1090, Austria.

*Supported by a National Institute of Justice Grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

†Co-first authors

Received 26 April 2016; and in revised form 23 Sept. 2016; accepted 9 Dec. 2016.

demands of age-at-death estimation in contemporary skeletal analysis and forensic identification contexts (1,21–26). We respond, specifically, to the recommendations for best practice in forensic case analysis given by the SWGANTH Age Estimation Advisory Committee: that age-at-death analyses should be carried out “with a spirit of scientific integrity” and “subjective interpretation should be limited, scientifically tested methods should be used, where possible, and method weaknesses and limitations should be communicated” (27).

We offer, therefore, improvement to age-estimation in forensic anthropology by developing a novel approach to quantifying bone surface morphology from laser scans. Our methods use three-dimensional coordinate data to permit a more thorough characterization of the age-related changes in pubic symphyseal shape and deliver a comprehensive framework for implementing a fully computational approach to the estimation of age-at-death. In this paper, we: 1) introduce a completely novel computational method that measures the curvature of the ventral margin of the scans and defines its relationship to the known age-at-death; 2) address the data limitations of our prior work, updating the proposed *SAH-Score* (28) and *TPS/BE algorithm* (29) methods, and report new models that incorporate additional data; and 3) combine these three aforementioned measures into two multivariate regression models that minimize the root-mean-square-error (RMSE) of the age-at-death estimate.

We present, for the first time, a standardized methodology for scanning the pubic bone surface, processing the three-dimensional meshes for coordinate data extraction, and implementing three complementary methods as individually executable techniques that merge shape analysis with age-estimation (the *SAH-Score*, *TPS/BE*, and *curvature*-based algorithms). Our results for all methods show a statistically significant relationship between the independent measures and the known ages-at-death. As each of these measures successfully, and differently, captures some of the age-associated changes of the symphyseal face morphology, we take advantage of their individual strengths by merging these methods together in multivariate procedures, attaining further improvement in results. Under this unified framework, we show how we can obtain a mathematically rigorous approximation of true chronological age that consistently outperforms the traditional pubic symphyseal scoring techniques by producing accurate and precise results, while simultaneously reducing the risk of method and observer subjectivity.

Materials and Methods

Data Acquisition

The models discussed here are built using 3D laser scans of the pubic symphysis, representing 93 adult individuals of forensic casework significance for whom sex, ancestry and true age-at-death are *a priori* known. The scans were taken with the NextEngine 3D Desktop Scanner, 2020i. Each pubic bone was positioned at 5 to 9 inches away from the scanner box with the symphyseal face being perpendicular to the base of the AutoDrive scanning stand. Each scan was produced in approximately 3.5 min using high definition scanning settings, specifying neutral image capture, 16 divisions, and 40K points per square inch. The scans were manipulated with the accompanying software, ScanStudio HD version 1.3.2, so that multiple scans were auto-aligned to form a single mesh. The mesh of the symphyseal face was isolated from the rest of the pubic bone, and the

surrounding areas of the bone were deleted, leaving only the trimmed surface mesh for the region of interest.

All 3D laser scanners capture the shape of an object by generating dense point clouds. The high-dimensional information from the scans can be stored in files of various formats. Our methods use PLY files (ASCII and binary), which simply consist of two lists of numbers. The first list gives the x , y , and z coordinates of each generated point (vertex) and the second list provides information on how the vertices are connected to form triangles. The angles at which the triangles' edges connect approximate the curvature of the shape. Figure 1 gives an example of a 3D scan available in our data, when shown as a seemingly smooth surface (top) and as a number of adjacent triangles (bottom).

Sample Properties

The true ages-at-death in our full dataset ($n = 93$) range from 16 to 90 years, with an equal number of pre-middle aged (younger than 40 years) and post-middle aged (older than 40 years) individuals. Figure 2 shows the distribution of ages present in our data. Studies have identified variation in pubic symphyseal morphology among different populations and between the sexes (6,7,12,16,30–34), therefore, we have limited our model building to data sourced from self-identified, or externally documented, white males. The majority of the scans, $n = 68$, in our data set were sourced from the W. M. Bass Donated Skeletal Collection, housed at the University of Tennessee, Knoxville (35). The Bass collection scans have dates of death that span the years 1987–2012. In addition, we included scans taken on two sets of publicly available casts, for which true ages are known and method-specific age-ranges are associated. Twelve casts are used for the application of the Suchey-Brooks system (5) and another 13 casts are used to implement McKern and Stewart's component-based scoring (3). By incorporating these cast-derived scans, we are able to evaluate the performance of our models against the well-established techniques for which these reference cast sets were originally produced. All Bass Collection scans represent the randomly selected left or right symphyseal surface of a unique individual. The six Suchey-Brooks phases are each represented by two different casts for a total of 12 Suchey-Brooks cast scans. Each of these casts was scanned and cleaned 3 times independently. These

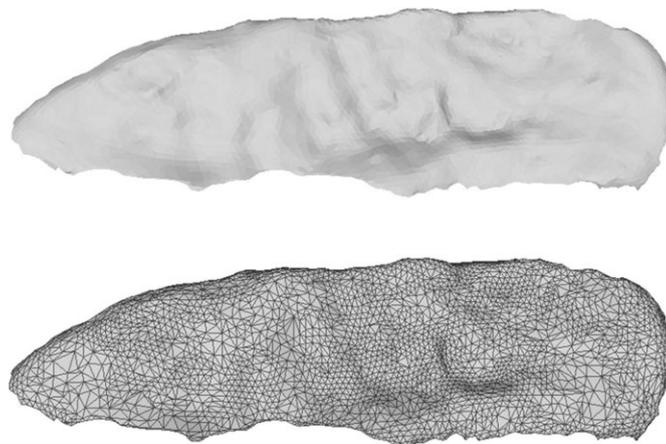


FIG. 1—An example of a 3D scan shown as a smooth surface (top) and as a number of adjacent triangles (bottom).

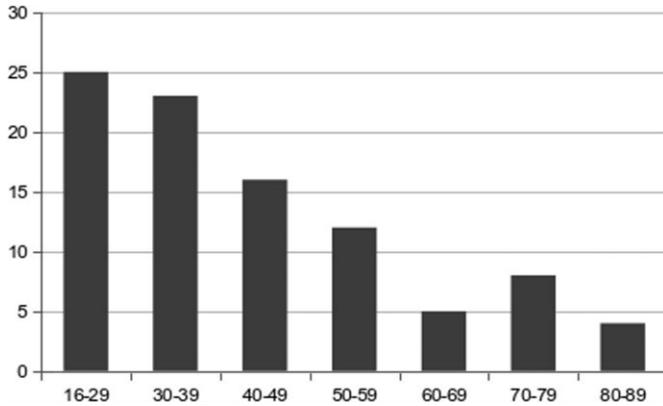


FIG. 2—Histogram of the age distribution of the data used in the presented models.

replicate scans were used in supplementary analyses to evaluate the impact that any variability in the scanning and cleaning processes may have on our estimated ages.

The original data used for the development of the *SAH-Score* and *TPS/BE* methods consisted of a subset of our currently available scans. The regression models presented in the *SAH-Score* paper (28) are built using 53 scans, while those given in the *TPS/BE* paper (29) make use of additional 3 scans from the Bass Collection, for a total of 56 scans. In both cases, the data was weighted by pubic symphyses that exhibit shape characteristics typical for older individuals, with less than 20% of the scans scored as Suchey-Brooks phases I, II, or III. The expectation is that with the addition of more data, thereby increasing quantity n and broadening the coverage of ages sampled, the results for both of the previously reported methods would improve. The new curvature method, as well as the combined-method regression models, both reported first here, were developed using the full complement of our available, 93, scans.

Surface & Outline Methods

This paper applies three computational shape analysis techniques to the morphology of the pubic symphysis, as captured by three-dimensional laser scans and represented by coordinate (x, y, z) data. These methods include:

- The *SAH-Score*, proposed by Slice and Algee-Hewitt (2015) that measures the variance on the symphyseal face to capture the gradual flattening of the surface associated with aging.
- The thin plate splines-based (*TPS*) measure proposed by Stoyanova, Algee-Hewitt and Slice (2015) that measures the bending energy (*BE*) required for transforming a perfectly flat, infinitely thin plate to match the surface of a pubic symphysis scan. Both the *SAH-Score* and *TPS/BE* methods capture the surface's gradual transition from ridged and furrowed to flat, but the *BE* accounts for the spatial structure of the analyzed shape.
- A newly developed component technique that measures the curvature of the ventral margin to capture the progressive formation of a rim around the entire symphyseal surface and its later erosion, accounting for the observation that older aged individuals present a depression of the symphyseal face relative to the rim.

Our three proposed computational methods aim at capturing age-related morphological features of the pubic symphysis that are recognized by previous studies and used in current practice

(2,3,5,14). Table 1 lists the specific morphology described in well-known, published studies and our computational method that quantifies it.

Data Manipulation

Prior to the use of any of this paper's three algorithms, the data needs to be standardized and some preprocessing steps need to be implemented, as described by Stoyanova, Algee-Hewitt and Slice (29). In order to standardize the scans' location, orientation and size, each scan is subjected to translation to the center of the coordinate system, rotation by the use of principal component analysis (PCA), and scaling to a uniform length. In addition, the scans' anatomical orientation is standardized so that the ventral, dorsal, superior and inferior halves are oriented in the same way. The *SAH-Score* algorithm makes use of all scan vertices but the *TPS* and the ventral curvature algorithms require the selection of a limited number of sample points that approximate the shape feature in question. The sample points used by the *TPS* algorithm are essentially a rectangular mesh covering the entire scan surface. Stoyanova, Algee-Hewitt and Slice (29) provide the computational details of selecting the vertices and discuss the use of meshes with different density.

The new algorithm introduced here for assessing the curvature of the ventral margin of the symphyseal face requires the selection of a number of equidistant semi-landmarks along the ventral margin of the scans. This is achieved by the same sampling grids used for the *TPS* algorithm. First, meshes with different densities are generated that cover the entire surface of the scans and later the inside points are identified and deleted leaving only vertices around the outline. While the *TPS/BE* paper (29) discusses the results of densities that vary in size from 11 points along the length to the scans to 101 points, the sparse grids, those with 11 and 21 vertices along the length, do not adequately account for the morphology of the symphyseal outline. Therefore, the results of using different numbers of outline semi-landmarks, ranging from 31 to 101, are compared in this study.

Curvature of the Ventral Margin

One of the changes that occur on the outline of the pubic symphysis is its progressive evolution from a more narrow shape (greater length to width ratio) towards a more rounded shape (smaller length to width ratio), a transition which is usually completed by Suchey-Brooks phase IV (5). We have developed, and present first here, a novel computational method that focuses

TABLE 1—Summarizes the specific age-related morphological changes captured by our computational methods and previous studies that discuss the relevant features.

Age-related Morphological Feature	Publication Describing the Feature	Our Proposed Method for Capturing the Feature
Transition of the symphyseal face from having a billowing surface (ridges and furrows; grooves) to a flat surface.	Todd, 1920 (2) McKern and Stewart, 1957 (3) Brooks and Suchey, 1990 (5) Milner and Boldsen, 2013 (14)	<i>SAH-Score</i> <i>TPS/BE Score</i>
Complete oval outline	Todd, 1920 (2) Brooks and Suchey, 1990 (5)	<i>Curvature of the ventral margin</i>

particularly on this age-related “outline” characteristic. The associated algorithm uses least-squares to find the best-fitting circle through the selected semi-landmarks on the ventral half of the symphyseal outline. It minimizes the distances between the semi-landmarks and the circle and returns its radius, r , as well as the x - y location of its center. The curvature of the ventral margin is measured as $1/r$. Figure 3 illustrates an example of the selected semi-landmarks along the ventral margin of one of the scans and the best-fitting circle through the vertices. A large radius implies a small curvature value (observed in younger individuals) and a smaller radius produces higher curvature value (typical for older individuals). This age-related shape characteristic is clearly illustrated in Figure 4, which shows the outlines of the scans of individuals whose ages-at-death are, respectively, from left to right, 25, 36, and 85 years. The outlines are oriented so that the ventral margin is to the right.

Algorithm

Given the set of data points (x_i, y_i) for $i = 1, 2, \dots, n$, we seek the circle that best fits the data. The equation for a circle with radius r and center located at (c_x, c_y) is $(x-c_x)^2 + (y-c_y)^2 = r^2$, which is equivalent to $a(x^2 + y^2) + bx + cy + d = 0$, where $a \neq 0$ is arbitrary and $b = -2c_x a, c = -2c_y a, d = -a(r^2 - c_x^2 - c_y^2)$. Thus, we seek to minimize

$$\min \sum_{i=1}^n (a(x_i^2 + y_i^2) + b x_i + c y_i + d)^2, \text{ subject to } a^2 + b^2 + c^2 + d^2 = 1.$$

Alternatively, the equation can be expressed in matrix form

$$\min \|Av\|,$$

where $v = (a, b, c, d)^T$ and A is the $n \times 4$ matrix

$$A = \begin{pmatrix} x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^2 + y_n^2 & x_n & y_n & 1 \end{pmatrix}.$$

The solution to $\min \|Av\|$ is given by the right singular vector of A associated with the smallest singular value. In other words,

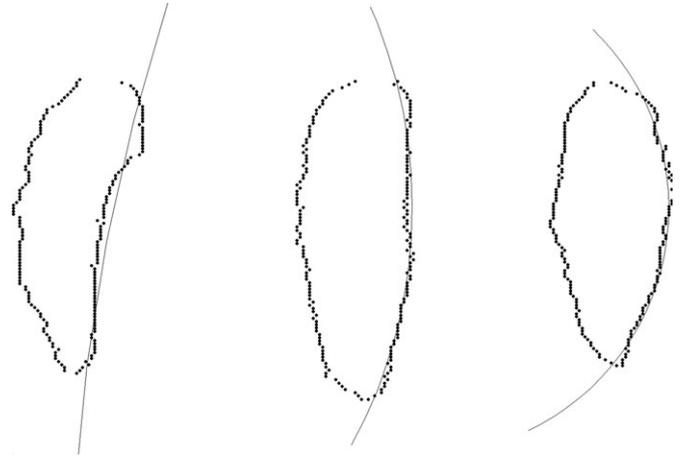


FIG. 4—Example of the selected semi-landmarks on three of the scan outlines and part of the best-fitting circle going through the vertices. The individuals have known ages-at-death equal to 25, 36, and 85 years (left to right).

we use the singular value decomposition $A = USV^T$ and take the last column of V . From that, $c_x = -b/2a, c_y = -c/2a$, and $r = \sqrt{c_x^2 + c_y^2 - d/a}$.

SAH-Score

The first fully quantitative and objective method for age-at-death estimation using 3D scans of the pubic symphysis was proposed by Slice and Algee-Hewitt (28). This *SAH-Score* method proceeds by subjecting the scan vertices to PCA and, then, by successively identifying the first principal component (PC) with the dimension with the greatest coordinate variance (the length of the scan), the second PC with the dimension with the second highest variance (the width of the scan) and the third PC with the dimension with the smallest variance (the depth of the scan). The eigenvalues associated with each of the PCs is proportional to the variance of the coordinates in the given dimension. The age-progressive transition of the symphyseal face - from being covered in ridges and furrows towards complete

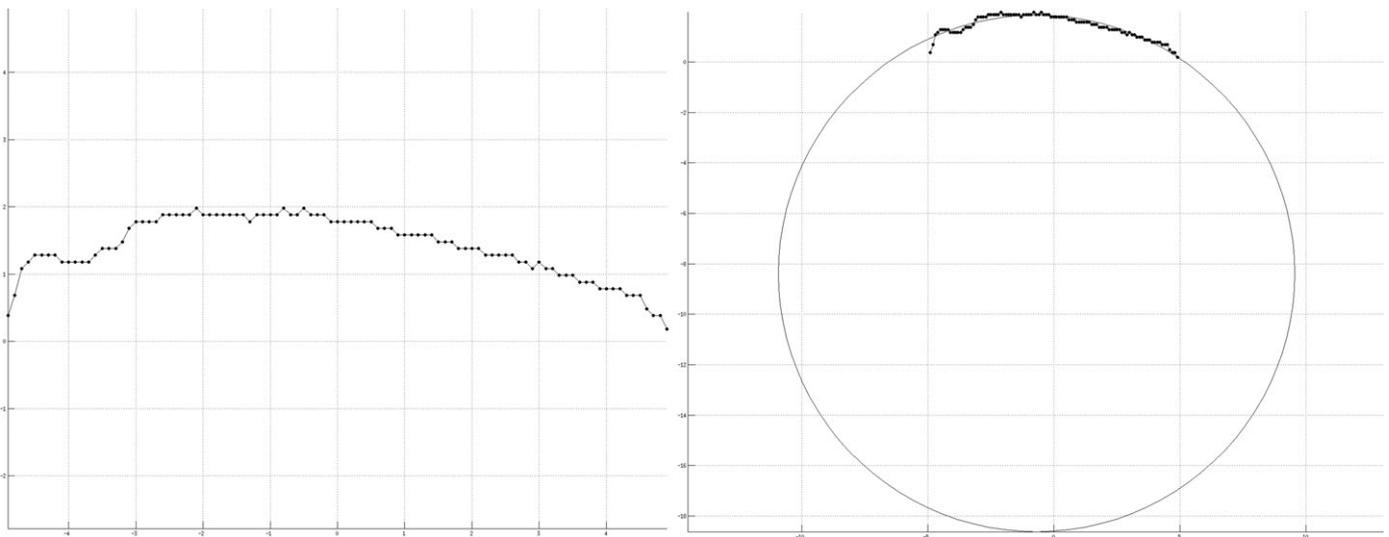


FIG. 3—The semi-landmarks generated over the ventral margin of one of the scans (right) and the best-fitting circle (left).

surface flattening - is captured by the change in variance in the third dimension. This method is shown, therefore, to produce high variance values for the younger individuals and low variance values for the older individuals (28).

Thin Plate Splines

The *TPS* algorithm was proposed as an alternative to the *SAH-Score* because, unlike a variance-based method, it accounts for the spatial structure of the shape (29): distinguishing, for instance, between the two different structures in Figure 5 by producing a higher *BE* value for the example on the left and a lower *BE* value for the example on the right. The algorithm simulates the bending of an infinitely thin, flat plate to match the scan surface and estimates the energy required for the bending. It uses two sets of corresponding control points: the first consists of the x , y , and z coordinates of the equidistant sample points selected on the scan surface, while the second represents the flat plane. The vertices in the second control set have the same x and y coordinates as the first set but all z coordinates are positioned at 0 so that the second set of control points lies in the x - y plane. Then, a linear system is devised using the vertex coordinates and the distances between them. The right-hand side of the system is a vector of the z coordinates of the vertices on the scan that are the targets. The solution of the linear system can be used to estimate the energy required for the z values on the plane to be transformed to the z values on the scan surface. This energy is taken as the measure for complexity of the scan surface and is shown to be associated with the exact age-at-death (29). The algorithm is discussed in detail in Stoyanova, Algee-Hewitt and Slice (29).

Results

Curvature of the Ventral Margin

A measure of the curvature of the ventral margin was calculated for all of the 93 pubic bone scans, representing self-identified white males with known ages-at-death. A linear regression analysis was performed on the relationship between the exact age-at-death and the value of the ventral curvature. The model uses the actual curvature value and log of the exact age-at-death. We are interested in minimizing the RMSE between the exact and estimated ages, defined as $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{estimated}_i - \text{exact}_i)^2}$, where n is the sample size. The results of the regression analysis (p -value, R^2 , and RMSE) of

using a different number of semi-landmarks along the outline are included in Table 2. The small variation in results for the different number of sample vertices indicates that the measure accounts for the general shape of the ventral margin rather than the details on the outline. The best results are obtained by the use of 61 sample points: p -value = 7.6×10^{-7} , R^2 value = 0.2365 and RMSE = 16.5457. The equation of the regression line is $y = 3.0739 + 6.7344x$, where y is $\log(\text{age})$ and x is the value of the ventral curvature. Figure 6 illustrates the results of the regression analysis. The Suchey-Brooks cast scans are shown as filled gray circles and the two scans that correspond to each phase are connected by a straight line, with the phase numbers shown by the connecting lines. The McKern and Stewart casts are similarly identified but by filled black circles.

Improvement of the SAH and TPS/BE Method Scores

The original *SAH-Score* publication reports results for a regression model that uses log of the known age-at-death and log of the surface variance, giving p -value = 0.0036, R^2 value = 0.1974, and RMSE = 17.76 years (28). The results previously reported for the *TPS/BE* method are also based on a log-log regression model, such that p -value = 0.0001, R^2 value = 0.2597, and RMSE = 18.3498 (29). With this paper's new introduction of approximately 40 additional scans the results for the *SAH-Score* regression model are revised to: p -value = 1.6×10^{-14} , R^2 value = 0.4786, and RMSE = 14.1511 years. For the *BE*, the similarly updated results are p -value = 4.7×10^{-9} , R^2 value = 0.3151, and RMSE = 16.3831 years. In both cases we demonstrate clearly that with the addition of samples there is a substantial improvement in model statistics and estimated values. In the case of the *SAH-Score* the R^2 value improvement is about 28% and the RMSE is decreased by 3.5 years. For the *TPS/BE* method, the R^2 value is increased by 5.5% and the RMSE is decreased by 2 years. The new p -values are both significant at the 0.05 level. The results are summarized in Table 3. The equations for the new regression lines for the *SAH-Score* and for the *BE* respectively are $y = 2.0595 - 0.8496x$ and $y = 3.6024 - 0.4069x$, where y is $\log(\text{age})$ and x is $\log(\text{SAH-Score})$ for the first equation and $\log(\text{BE})$ for the second equation. Figures 7 and 8 show the new regression models for the *SAH-Score* and *TPS/BE* methods respectively and, similar to Figure 6, the Suchey-Brooks cast scans are represented by filled gray circles connected by straight lines, the McKern and Stewart casts are shown as filled black circles. In the case of the *BE*, the old and new results are consistent in that the same sample grid size (21 vertices along the length of the scans) produces best results.

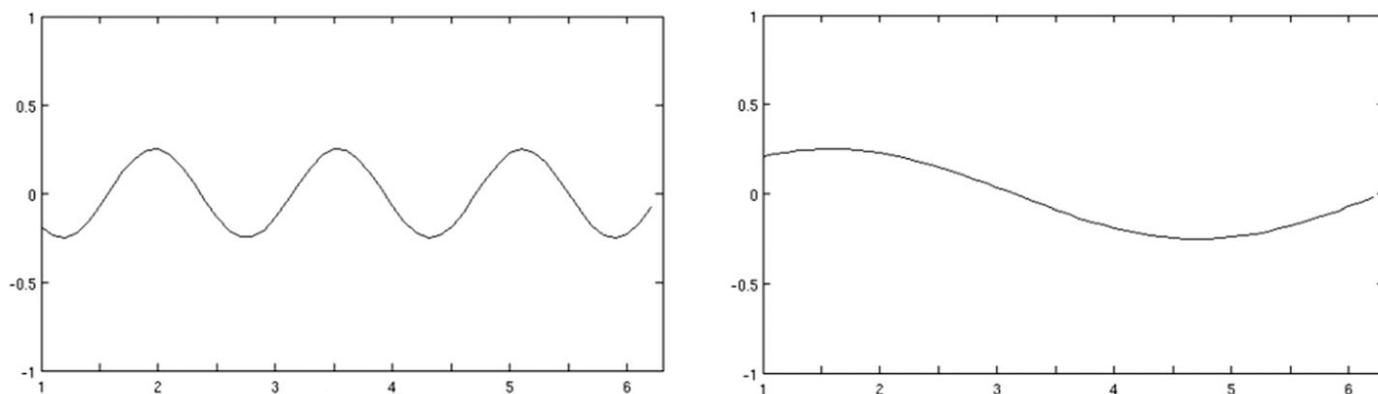


FIG. 5—Examples of structures that will produce the same *SAH-Score* (variance) but different bending energy values.

TABLE 2—Results of using the curvature of the ventral margin of the outline for the different sample grid sizes.

Grid Size	p-Value	R ²	RMSE
31	9.6×10^{-6}	0.1946	16.9702
41	5.8×10^{-6}	0.2031	16.8586
51	1.0×10^{-6}	0.2319	16.5756
61	7.6×10^{-7}	0.2365	16.5457
71	1.1×10^{-6}	0.2305	16.6891
81	9.5×10^{-6}	0.2330	16.6096
91	8.6×10^{-6}	0.2347	16.6459
101	1.8×10^{-6}	0.2222	16.7431

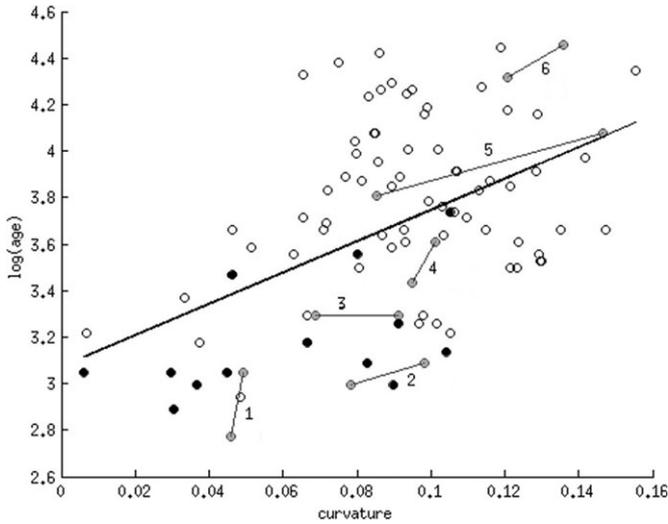


FIG. 6—The regression model for the ventral curvature margin. The Suchey-Brooks casts are shown as filled gray circles, and the two casts in each phase are connected by a straight line. The corresponding phase numbers are shown by the connecting lines. The McKern and Stewart casts are shown as filled black circles.

Combining the Measures

Each of the three measures demonstrates a strong relationship with the known age-at-death. The two surface scores both measure the flatness of the symphyseal face and therefore capture similar shape characteristics. The curvature of the ventral margin provides unique shape information that is unaccounted for by the surface measures. Therefore, it is reasonable to expect that some combination of surface and outline measures will improve the age-estimation, in accuracy and precision or both. Table 4 shows the correlation coefficients of the three measures. The results of combining the SAH-Score and the curvature of the ventral margin in a multivariate regression model are p -value = 6.0×10^{-15} , adjusted $R^2 = 0.5178$, and RMSE = 13.6830 years. The observed improvement is 4% in the R^2 and

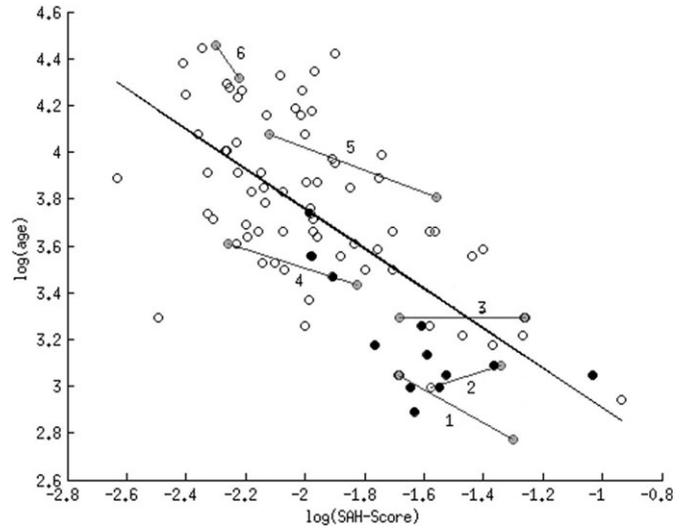


FIG. 7—The log-log regression model for the SAH-Score using the new data. The Suchey-Brooks casts are shown as filled gray circles, and the two casts in each phase are connected by a straight line. The corresponding phase numbers are shown by the connecting lines. The McKern and Stewart casts are shown as filled black circles.

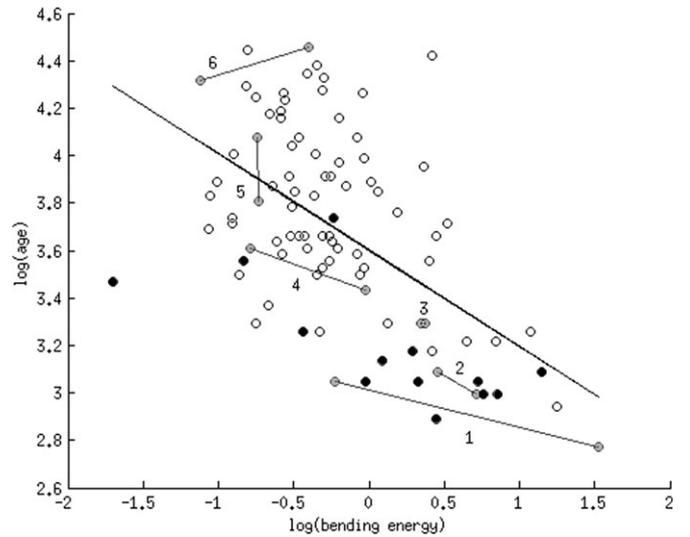


FIG. 8—The log-log regression model for the bending energy using the new data. The Suchey-Brooks casts are shown as filled gray circles, and the two casts in each phase are connected by a straight line. The corresponding phase numbers are shown by the connecting lines. The McKern and Stewart casts are shown as filled black circles.

about a half year in the RMSE. The results of combining the TPS/BE method with the curvature of the ventral margin produces a p -value = 1.3×10^{-11} , adjusted $R^2 = 0.4267$, and

TABLE 3—Summary of the results of the SAH-Score and the TPS/BE methods applied to both the old and the new datasets and the improvement offered by the additional data.

	SAH-Score			TPS/BE Method		
	p-Value	R ²	RMSE	p-Value	R ²	RMSE
Old data	0.0036	0.1974	17.76	0.0001	0.2597	18.3498
New data	1.6×10^{-14}	0.4786	14.1511	4.7×10^{-9}	0.3151	16.3831
Improvement		0.2812	-3.6089		0.0554	-1.9667

TABLE 4—The correlation coefficients of the three measures.

Measure	SAH-Score	BE	Ventral Curvature
SAH-Score	1.0000	0.5594	-0.4467
BE	0.5594	1.0000	-0.2351
Ventral Curvature	-0.4467	-0.2351	1.0000

RMSE = 15.0704 years. The improvement is 11% in R^2 value and 1.3 years in RMSE. A summary of the results is presented in Table 5. The equations of the regression lines are $y = 2.015 + 3.0674a - 0.7281b$ and $y = 3.1805 + 4.8460a - 0.3312c$, where y is $\log(\text{age})$, a is the value of the ventral curvature, b is the $\log(\text{SAH-Score})$ and c the $\log(\text{BE})$. Figures 9 and 10 show the known age-at-death versus the estimated age-at-death for the two multivariate regression models.

Evaluation of the Models

All models were subjected to cross-validation analysis, where the data set is split into two subsets with equal number of scans with the same age distribution as the original set. A model is built using each one of the subsets and the ages of the excluded scans are approximated by that model. The cross-validation RMSEs for the SAH-Score combined with the curvature value for the two sets are 14.31 and 12.96 years, while the cross-validation RMSEs for the TPS/BE and curvature measures are 15.49 and 15.01 years respectively.

We estimated the difference between the exact age-at-death and the estimated age for each scan using each of the 5 regression models. The average age-distances for the ventral curvature, the SAH-Score and the TPS/BE method respectively are 12.8, 10.9, and 12.7 years. The corresponding numbers for the two multivariate regression models are 10.8 years (ventral curvature + SAH-Score) and 11.4 years (ventral curvature + TPS/BE). The average among the 5 models is 11.7 years. In addition, our results indicate that 28–34% of the age-estimates are within 5 years of the known age-at-death, 17–32% are within 5–10 years of the exact age and 8.6–19% are within 10–15 years. Overall, more than half of the estimates are within 10 years of the exact age-at-death and 64.5–75% are within 15 years. Tables 6 and 7 present the aforementioned results in detail.

For comparison against the performance of the Suchey-Brooks method, Table 8 shows the exact ages of the Suchey-Brooks casts, their corresponding phase with associated age-interval, and their estimated ages using the SAH-Score, the TPS/BE, and both multivariate models. It can be observed from the table that with the exception of the second phase I scan all of our age-estimates are within the associated age-interval. In addition, the age-intervals produced by our models (2xRMSE) range from about 27 to 32 years and are generally smaller than the Suchey-Brooks phase intervals.

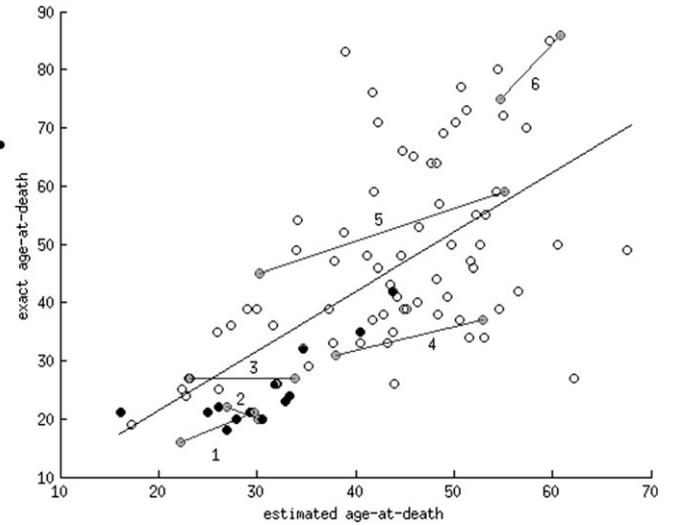


FIG. 9—The exact age-at-death versus the age-at-death estimated by a multivariate regression model that includes the SAH-Score and the curvature of the ventral margin. The Suchey-Brooks casts are shown as filled gray circles, and the two casts in each phase are connected by a straight line. The corresponding phase numbers are shown by the connecting lines. The McKern and Stewart casts are shown as filled black circles.

Testing Repeatability

The models presented so far include the second set of the replicate Suchey-Brooks cast scans. Additional models are built using the first and the third replicate sets. Table 9 summarizes the results of the three sets of cast scans. Further, it is important to evaluate whether the replicating scans produce similar age-estimates. The ages of the casts in sets 1 and 3 are estimated by the model that includes the scans in replicate set 2. On average the difference between the estimated ages of set 2 and the estimated ages of set 1 for the TPS/BE combined with the curvature value is 4.16 years, the difference between set 2 and set 3 is 2.75 years. For the SAH-Score and the curvature value, the respective values are slightly less at 3.91 and 1.5 years.

Additional Model Performance Tests

We evaluate the performance of our methods against recent publications (36–39) by calculating the bias and inaccuracy values (Tables 10 and 11) for each of our proposed models as well as the mean of the age estimated and the correlation coefficient between the exact ages-at-death and the estimated ages-at-death (Table 12). The bias and inaccuracy values are calculated using the formulae: $\text{Bias} = \sum (\text{estimated age} - \text{exact age})/N$ and $\text{Inaccuracy} = \sum |\text{estimated age} - \text{exact age}|/N$, where N is the sample size. The bias shows whether the individuals in a

TABLE 5—Summary of the results of the SAH-Score and the BE used in univariate regression models versus the multivariate results when combined with the curvature value and the improvement.

	SAH-Score			TPS Method		
	p-Value	R ²	RMSE	p-Value	R ²	RMSE
Univariate	1.6×10^{-14}	0.4786	14.1511	4.7×10^{-9}	0.3151	16.3831
Multivariate	6.0×10^{-15}	0.5178	13.6830	1.3×10^{-11}	0.4267	15.0704
Improvement		0.0392	-0.4681		0.1116	1.3127

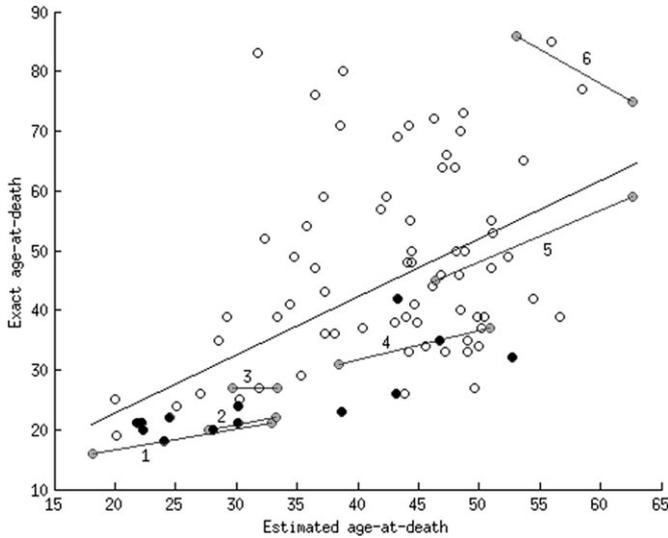


FIG. 10—The exact age-at-death versus the age-at-death estimated by a multivariate regression model that includes the bending energy and the curvature of the ventral margin. The Suchey-Brooks casts are shown as filled gray circles, and the two casts in each phase are connected by a straight line. The corresponding phase numbers are shown by the connecting lines. The McKern and Stewart casts are shown as filled black circles.

TABLE 6—For each method, the percent of age-estimates that fall within 5, 10, and 15 years of the exact age-at-death of the individuals.

Difference in Years	SAH-Score	TPS/BE	Ventral Curvature	SAH-Score + Outline	BE + Outline
≤ 5	34.4%	27.9%	30.1%	32.2%	34.4%
6–10	21.5	25.8	18.2	32.2	17.2
11–15	19.3	13.9	16.1	8.6	19.3
> 15	24.7	32.2	35.4	26.8	29.0

given age group are consistently overestimated or underestimated and the inaccuracy is a measure of the error similar to the RMSE. Table 10 shows that, for each of the five models, the males who are between 40 and 49 years old are aged with the least bias (as low as 0.06) and lowest inaccuracy (between 5.31 and 7.93 years). Table 12 demonstrates that the correlations

between the exact ages-at-death and the estimated ages-at-death are robustly positive between 0.41 and 0.65 (p -value < 0.05), suggesting a monotonic relationship with high degree of linear dependence. The means of the estimated age-at-death for the five models are within 2 years of the sample mean.

Discussion

Reinforcing our prior work, with the addition of a novel curvature-based method and the refinement of our previous methods through expanded case sampling, we demonstrate that objective and reliable age-at-death estimation can be obtained by the use of computational methods and 3D laser scans. The results presented here show a significant improvement over those initially reported in the *SAH-Score* (28) and *TPS* (29) papers. Grand means calculated across all three methods show how we achieve estimates for 29.2% of our sample that fall within a 5 year range, 50.6% for a 10 year range, and 66.1% for a 15 year range of their true age. When combined, we arrive at even better results: increasing these values to 33.3%, 58%, and 72% respectively. It can be argued that, by attaining a RMSE as low as 13.68 years, we offer objective computational methods that outperform the current gold-standard, Suchey-Brooks, age-estimation technique.

We demonstrate that further improvement can be achieved by incorporating additional shape measures that capture unique information. To this end, we have proposed an outline-based score that measures, specifically, the curvature of the ventral margin. As a single measure, the curvature of the ventral margin does not outperform either one of the surface measures, however, it adds improvement to both the *SAH-Score* and *TPS/BE* methods when combined in multivariate regression models. The proposed outline score addresses a specific concern about the accurate representation of the margin of the pubic symphysis by the scan data. In particular, the symphyseal face needs to be manually selected from the rest of the bone. The cleaning process depends to some extent on the resolution of the scan data. Additionally, in younger individuals there is a lack of delimitation of the extremities (5). Studying the outline morphology in great detail may introduce error in the calculations. Therefore, we expect that methods that capture the overall shape of the outline can provide more reliable information. The proposed

TABLE 7—For each method and for each age group the percent of age-estimates that fall within 5, 10, and 15 years of the exact age-at-death of the individuals.

Age-at-death	Difference in Years	SAH-Score	TPS/BE	Ventral Curvature	SAH-Score + Outline	BE + Outline
<30	≤ 5	44%	32%	20%	40%	48%
	6–10	24	32	28	48	28
	11–15	24	12	20	4	8
	>15	8	24	32	8	16
30–39	≤ 5	30.4	43.5	47.8	30.5	21.8
	6–10	43.5	39.1	17.4	47.8	21.8
	11–15	21.7	4.3	8.7	4.3	39.1
	>15	4.3	6.3	26.1	17.4	17.3
40–49	≤ 5	50	37.5	56.25	43.75	62.5
	6–10	12.5	25	25	31.25	18.75
	11–15	25	31.25	18.75	18.75	18.75
	>15	12.5	6.25	0	6.25	0
50–59	≤ 5	50	16.7	25	50	41.7
	6–10	8.3	25	16.7	16.7	8.3
	11–15	25	33.3	25	16.7	16.7
	>15	16.7	25	33.3	16.7	33.3
>60	≤ 5	0	0	0	0	0
	6–10	5.9	0	0	0	0
	11–15	0	0	11.8	5.9	11.8
	>15	94.1	100	88.2	94.1	88.2

TABLE 8—Results for the 12 Suchey-Brooks casts. The table reports the exact age-at-death, Suchey-Brooks phase and associated age-interval for each cast as well as the age-estimates produced by the surface measures as univariate models and the two multivariate regression models.

Suchey-Brooks Phase	Phase Age-range	Exact Age	SAH-Score	BE	SAH + Outline Estimate	BE + Outline Estimate
I-I	15–23	16	24	20	22	18
I-II		21	33	40	30	33
II-I	19–34	22	25	30	27	33
II-II		20	30	27	30	28
III-I	21–46	27	33	32	34	33
III-II		27	23	31	23	30
IV-I	23–57	31	37	37	38	38
IV-II		37	53	51	53	51
V-I	27–66	59	48	50	55	63
V-II		45	29	49	30	46
VI-I	34–86	75	52	58	55	63
VI-II		86	55	43	61	53

TABLE 9—Results of the combined models for the replicating sets of the Suchey-Brooks cast scans.

	SAH-Score + Outline			BE + Outline		
	p-Value	R ²	RMSE	p-Value	R ²	RMSE
Set1	9.0 × 10 ⁻¹⁵	0.5129	13.8630	1.4 × 10 ⁻¹¹	0.04257	15.1522
Set2	6.0 × 10 ⁻¹⁵	0.5178	13.6830	1.3 × 10 ⁻¹¹	0.4140	15.0704
Set3	2.5 × 10 ⁻¹⁴	0.5012	13.8393	5.0 × 10 ⁻¹¹	0.4094	15.2896

measure for the curvature of the ventral margin addresses this requirement by demonstrating low dependence on the number of selected semi-landmarks.

When used in univariate regression models against the known age-at-death, the SAH-Score outperforms the BE by more than 15% in R² value and 2.2 years in RMSE. When combined with the outline measure the models that use the SAH-Score continue to perform better, however, the results of the multivariate regression models are more similar. The difference between the adjusted R² values is 10% and the difference of the RMSEs is about 1.5 years. This indicates that some of the outline information that was unaccounted for by the TPS/BE was actually captured by the SAH-Score.

The multivariate regression models were subjected to cross-validation analysis where a model is built using only half of the data and is used to evaluate the excluded scans. The two estimated cross-validation RMSEs for both methods are within less than a year from the RMSEs of using all the data, which attests for the models’ accuracy and stability in practice. Further, the two additional scans of each of the Suchey-Brooks casts were used for testing purposes to show how the scanning and cleaning process affect the results of the models. Although not identical, the three sets of scans produce, overall, comparable age-estimates.

Tests of capability between age-estimation approaches, including very new techniques, are critical in support of method utility and for highlighting areas that need or have undergone improvement (36–39). Accordingly, we have generated bias and inaccuracy estimates for our methods (Tables 10 and 11), as well as correlation coefficients (Table 12) that serve as comparative data between our results and those obtained with the Suchey-Brooks method. The bias values show that our methods consistently overestimate the ages of young individuals (< 40) and underestimate the ages of very old individuals (> 60). When comparing our inaccuracy values to Miranker (36), we observe that our models perform least well for the 20–40 age group. Importantly, however, we show that they are more accurate when applied to

TABLE 10—The bias and inaccuracy values calculated for each of our five models. The results are reported by age group as well as for the entire dataset.

Age-at-death	SAH-Score		BE		Ventral Curv		SAH + Outline		BE + Outline	
	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias
< 30	8.32	7.2	9.12	8.96	11.4	11.08	7.64	6.12	7.04	6.64
30–39	7.69	3.95	8.39	7.34	8.60	6.69	8.91	5.17	10.21	8.30
40–49	7.93	1.43	7.56	0.06	5.18	-3.93	7.62	1	5.31	-0.43
50–59	7.58	-6.08	12.08	-12.08	11.25	-10.58	7.75	-5.41	10.16	-9.5
≥ 60	23.76	-23.76	28.41	-28.41	29.11	-29.11	23.11	-23.11	26	-26
Entire dataset	10.81	-1.96	12.58	-2.51	12.86	-2.73	10.79	-1.82	11.39	-2.21

TABLE 11—The bias and inaccuracy values calculated for each of our five models is given. For ease of comparison, the results are reported using the same age groups as in Miranker (36).

Exact age-at-death	SAH-Score		BE		Ventral Curv		SAH + Outline		BE + Outline	
	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias
20–40	8.1	5.71	9.08	8.47	9.89	5.84	8.36	5.76	8.93	7.76
41–60	7.66	-2.25	9.22	-5.96	7.88	-6.85	7.74	-2.03	7.33	-4.81
61–80	21.57	-21.57	25.28	-25.28	27.28	-27.28	21.35	-21.35	23.5	-23.5
≥ 81	34	-34	43	-43	37.66	-37.66	31.33	-31.33	37.66	-37.66

TABLE 12—The mean values of the age-estimates for each computational model and the Pearson correlation between the exact ages-at-death and the estimated ages-at-death is presented.

Exact age-at-death mean	SAH-Score		TPS/BE		Ventral Curvature		SAH-Score + Outline		BE + Outline	
	Mean	Correlation	Mean	Correlation	Mean	Correlation	Mean	Correlation	Mean	Correlation
43.25	41.29	0.6225	40.74	0.4339	40.52	0.4128	41.43	0.6531	41.04	0.5541

all other age groups. Only minor exceptions are when the *BE/TPS* score is used as a single measure for individuals older than 81 years and the curvature of the ventral margin is used as a single measure for individuals older than 60 years. It should be noted that, for the 41–60 age group, the error from our multivariate regression models is half of that reported by Miranker (36). The correlation coefficients between the exact ages-at-death and the estimated ages-at-death are consistent with Merritt's findings: "There are moderate, positive, significant correlations for... the Suchey-Brooks (0.569, $p < 0.001$) ... mean ages-at-death to the actual ages-at-death." (37).

What is important to learn from these comparisons is that our methods offer improvement in age-estimation for some cohorts and offer improvement in objectivity across all estimates. The methods given here respond, therefore, to the increasing expectations of rigor in the forensic sciences. Specifically, they can help skeletal biologists overcome the emerging issues surrounding criteria of "reasonable scientific certainty," recalling recent statements made by the Department of Justice against the suitability of these kind of evidentiary standards (40).

Software

Multiplatform, open source software called forAGE is available for download at (<http://morphlab.sc.fsu.edu/>). Example scan files of a subset of the standard comparative casts are included with the program to allow all forensic practitioners to test the software and compare results from their own scans of these same casts. Currently, the program accepts a PLY scan file as input, calculates the *SAH-Score* and the *TPS/BE* and estimates the age-at-death using the regression models presented in the surface analysis papers (28,29). To accompany this new publication, we have produced an updated version of forAGE for immediate release that also calculates the curvature of the ventral margin and estimates the age-at-death using the multivariate regression models presented here.

Acknowledgments

We thank our colleagues at the University of Tennessee for granting access to the skeletal materials and scanning resources that facilitated this work. We also express our deepest appreciation to the donors and their families who have contributed to the W. M. Bass Collection - without their support for the advancement of science through skeletal research, studies such as this would not be possible.

References

1. Algee-Hewitt BFB. Age estimation in modern forensic anthropology. In: Tersigni-Tarrant MT, Shirley NR, editors. *Forensic anthropology: an introduction*. Boca Raton, FL: CRC Press, 2013;181–230.
2. Todd TW. Age changes in the pubic bone. *Am J Phys Anthropol* 1920;4:285–327.
3. McKern TW, Stewart TD. Skeletal age changes in young American males analysed from the standpoint of age identification. Technical report

- EP-45. Natick, MA: Quartermaster Research and Development Command, 1957.
4. Brooks ST. Skeletal age at death: the reliability of cranial and pubic age indicators. *Am J Phys Anthropol* 1955;13:567–97.
5. Brooks S, Suchey JM. Skeletal age determination based on the os pubis: a comparison of the Acsadi-Nemeskeri and Suchey-Brooks methods. *Hum Evol* 1990;5:227–38.
6. Katz D, Suchey JM. Age determination of the male OS pubis. *Am J Phys Anthropol* 1986;69:427–35.
7. Katz D, Suchey JM. Race differences in pubic symphyseal aging patterns in the male. *Am J Phys Anthropol* 1989;80:167–72.
8. Milner GR, Boldsen JL. Transition analysis: a validation study with known-age modern American skeletons. *Am J Phys Anthropol* 2012;148:98–110.
9. Buikstra JE, Ubelaker DH. Standards for data collection from human skeletal remains: proceedings of a seminar at the field museum of natural history. Fayetteville: Arkansas Archeological Survey, 1994.
10. Meindl RS, Lovejoy CO, Mensforth RP, Walker RA. Method of age determination using the OS pubis, with a review and tests of other current methods of pubic symphyseal aging. *Am J Phys Anthropol* 1985;68:29–45.
11. Garvin HM, Passalacqua NV. Current practices by forensic anthropologists in adult skeletal age estimation. *J Forensic Sci* 2012;57:427–33.
12. Berg GE. Pubic bone age estimation in adult women. *J Forensic Sci* 2008;53:569–77.
13. Hartnett KM. Analysis of age-at-death estimation using data from a new, modern autopsy sample - part I: pubic bone. *J Forensic Sci* 2010;55:1145–51.
14. Milner GR, Boldsen JL. Transition analysis age estimation: skeletal scoring manual. *Fordisc Version 1.00*, 2013; <http://anth.la.psu.edu/research/bioarch/docs/TAManual2013June.pdf> (accessed September 15, 2016).
15. Konigsberg LW, Frankenberg SR, Liversidge HM. Optimal trait scoring for age estimation. *Am J Phys Anthropol* 2015;159:557–76.
16. Kimmerle EH, Jantz RL, Konigsberg LW, Baraybar JP. Analysis of age-at-death estimation through the use of pubic symphyseal data. *J Forensic Sci* 2008;53:558–68.
17. Klepinger LL, Katz D, Micozzi MS, Carroll L. Evaluation of cast methods for estimating age from the os pubis. *J Forensic Sci* 1992;37:763–70.
18. Kimmerle EH, Prince DA, Berg GE. Inter-observer variation in methodologies involving the pubic symphysis, sterna ribs, and teeth. *J Forensic Sci* 2008;53:594–600.
19. Baccino E, Ubelaker DH, Hayek LA, Zerilli A. Evaluation of seven methods of estimating age at death from mature human skeletal remains. *J Forensic Sci* 1999;44:931–6.
20. Shirley NR, Ramirez Montes PA. Age estimation in forensic anthropology: quantification of observer error in phase versus component-based methods. *J Forensic Sci* 2015;60(1):107–11.
21. Konigsberg L, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 2008;53:541–57.
22. Baccino ED, Schmitt A. Determination of adult age at death in forensic context. In: Schmitt A, Cunha E, Pinheiro J, editors. *Forensic anthropology and medicine: complementary sciences from recovery to cause of death*. Totowa, NJ: Humana Press, 2006;259–80.
23. Milner GR, Wood JW, Boldsen JL. Advances in paleodemography. In: Katzenberg MA, Saunders SR, editors. *Biological anthropology of the human skeleton*, 2nd edn. New York, NY: Wiley-Liss, 2008;561–600.
24. Ritz-Timme S, Cattaneo C, Collins MJ. Age estimation: the state of the art in relation to the specific demands of forensic practice. *Int J Legal Med* 2000;113:129–36.
25. Franklin D. Forensic age estimation in human skeletal remains: current concepts and future directions. *Leg Med* 2010;12:1–7.
26. Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, et al. The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 2009;193:1–13.

27. Scientific Working Group for Forensic Anthropology (SWGANTH). Age estimation; <http://swganth.startlogic.com/Age%20Rev1.pdf> (accessed September 15, 2016).
28. Slice DE, Algee-Hewitt BFB. Modeling bone surface morphology: a fully quantitative method for age-at-death estimation using the pubic symphysis. *J Forensic Sci* 2015;60:835–43.
29. Stoyanova D, Algee-Hewitt BFB, Slice DE. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am J Phys Anthropol* 2015;158:431–40.
30. Gilbert BM, McKern TW. A method for aging the female OS pubis. *Am J Phys Anthropol* 1973;38:31–8.
31. Hoppa RD. Population variation in osteological aging criteria: an example from the pubic symphysis. *Am J Phys Anthropol* 2000;111:185–91.
32. Komar D. Lessons from Srebrenica: the contributions and limitations of physical anthropology in identifying victims of war crimes. *J Forensic Sci* 2003;48:713–6.
33. Lottering N, Macgregor DM, Meredith M, Alston CL, Gregory LS. 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* 2013;150:386–99.
34. Schmitt A. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. *Int J Osteoarchaeol* 2004;14:1–6.
35. Bass Collection, Forensic Anthropology Center and Dept of Anthropology, University of Tennessee at Knoxville, <http://fac.utk.edu/> (accessed September 23, 2016).
36. Miranker M. A comparison of different age estimation methods of the adult pelvis. *J Forensic Sci* 2016;61:1173–9.
37. Merritt CE. A test of Hartnett's revisions to the pubic symphysis and fourth rib methods on a modern sample. *J Forensic Sci* 2014;59:703–11.
38. Fleischman JM. A comparative assessment of the Chen et al. and Suchey-Brooks pubic aging methods on a North American sample. *J Forensic Sci* 2013;58:311–23.
39. Lungmus EK. An examination of error in the application of pubic aging techniques [thesis]. Missoula, MT: University of Montana, 2009.
40. Augenstein S. DOJ's code of conduct: no more 'reasonable scientific certainty'. *Forensic Mag* 2016 Sept 14; <http://www.forensicmag.com/article/2016/09/dojs-code-conduct-no-more-reasonable-scientific-certainty> (accessed September 23, 2016).

Additional information and reprint requests:
 Bridget F. B. Algee-Hewitt, Ph.D.
 Department of Biology
 Stanford University
 Stanford, CA 94305
 E-mails: bridgeta@stanford.edu; balgeehe@vols.utk.edu

TECHNICAL NOTE**ANTHROPOLOGY**

Detelina K. Stoyanova,¹ Ph.D.; Bridget F. B. Algee-Hewitt,^{1,2} Ph.D.; Jieun Kim ¹ Ph.D.; and Dennis E. Slice,^{1,3} Ph.D.

A Study on the Asymmetry of the Human Left and Right Pubic Symphyseal Surfaces Using High-Definition Data Capture and Computational Shape Methods*

ABSTRACT: The pubic symphysis is among the most commonly used bilateral age indicators. Because of potential differences between right and left sides, it is necessary to investigate within-individual asymmetry, which can inflate age estimation error. This study uses 3D laser scans of paired pubic symphyses for 88 documented White males. Scan data are analyzed by numerical shape algorithms, proposed as an alternative to traditional visual assessment techniques. Results are used to quantify the within-individual asymmetry, evaluating if one side produces a better age-estimate. Relationships between the asymmetry and advanced age, weight, and stature are examined. This analysis indicates that the computational, shape-based techniques are robust to asymmetry (>80% of paired differences are within 10 years and >90% are within 15 years). For notably more asymmetric cases, differences in estimates are not associated with life history factors. Based on this study, either side can be used for age-at-death estimation by the computational methods.

KEYWORDS: forensic science, age-at-death estimation, pubic symphysis, asymmetry, 3D laser scans, computational algorithms, morphometrics

For forensic anthropologists, age-at-death is among the most important of the demographic parameters that are subject to estimation from osteological indicators in the adult skeleton (1). In this forensic setting, knowing age-at-death for a set of unidentified human remains provides critical personal information that, when combined with other osteological data, may assist the medicolegal community in missing person, mass fatality, human rights, and violent crime casework (2). Therefore, the accuracy and precision with which age-at-death can be inferred, has the potential to directly affect whether the decedent is positively identified. In forensic literature, accuracy can be defined in two ways: (i) the distance between an estimated value and true value (3); or (ii) the ability of a prediction interval/range to capture the true value (4,5). On the other hand, precision can indicate two different properties of a method: (i) the repeatability of a method between-/within-observer at different times regardless of the distance between the estimated value and the true value (3); or (ii) the width of a predicted interval/range (4,5). In this study,

accuracy and precision follow the former definitions and, therefore, signify the distance between the age-estimate and documented chronological age and the repeatability of age estimation methods, respectively, when the methods are applied to an age indicator exhibiting varying degrees of bilateral asymmetry. Owing to the importance of age, there has been considerable focus placed on assessing skeletal indicators of age and method performance (4,6–23). Particular interest has been paid to the pubic symphysis and to the Suchey–Brooks (24–26) system, a method for the visual assessment of the symphysis and the estimation of age using discrete phases associated with known ages. This emphasis is not surprising as the pubic symphysis is historically one of the first skeletal features to be described in the context of age estimation (24,27–32) and continues to be the most widely studied age-progressive trait in the adult skeleton for both past and present populations (17,33–46). Moreover, today, the pubic symphysis and the Suchey–Brooks system are said to be the preferred skeletal indicator/method for age-at-death estimation by forensic anthropologists (10).

Asymmetry in the pubic bone morphology is a critical and potentially confounding factor to consider in age estimation, especially for cases where the pubic symphysis is the only available/preferred source for age prediction (17,47–49). It increases variation in the indicator's assessment, which may in turn contribute to age estimation error. Asymmetry here refers to differences in the feature morphology between the paired elements for the same age indicator, and specifically the pubic symphyseal surface. In this context, the left and right sides of the pubic symphysis may vary in appearance and in their assigned phase or

¹Department of Scientific Computing, Florida State University, Tallahassee, FL 32306.

²Department of Biology, Stanford University, Stanford, CA 94305.

³Department of Anthropology, University of Vienna, Vienna 1090, Austria.

Corresponding author: Detelina Stoyanova, Ph.D. E-mail: detelinastoyanova@gmail.com

*Supported by a National Institute of Justice Grant (2015-DN-BX-K010) awarded to the senior authors, Slice and Algee-Hewitt.

Received 15 Feb. 2018; and in revised form 29 April 2018, 11 June 2018, 21 June 2018; accepted 21 June 2018.

score (e.g., Todd, McKern and Stewart, Suchey–Brooks). If so, the estimated age (or the age range) of the individual may differ according to which side of the pelvis is subjected to review. Consequently, asymmetric features may compromise the accuracy of the biological profile in terms of the information given on age when the practitioner derives a final age-estimate, whether voluntarily or unavoidably (due to the condition of the remains), without consulting any other age indicators than the pubic symphysis. This risk is highest when there is no clear indication of which one of the sides, the left or right pubic symphyseal surface, displays the more “truthful” information on biological age for the given skeleton.

Asymmetry in phase classification occurs in nontrivial proportions, even exceeding 60% of cases in recent studies (17,47–49). As others have already cautioned, in the presence of asymmetry, there are important statistical effects on age estimation as accuracy drops significantly at a reported rate of $\approx 20\%$ (47). There are also practical consequences. For example, any apparently miss-matched bones found outside of clear paired, single individual context (e.g., when commingling is suspect as in mass fatality scenes) could be treated independently. They would be assigned to different phases and age intervals (48), and their discordant age-related morphologies could lead to an erroneous accounting of the minimum number of individuals (MNI) at the recovery site (49).

Some strategies for estimating the age-at-death in the presence of more notable asymmetry include taking the mean age of the two sides as suggested by Schmitt, 2004 (48). Schmitt’s results show that calculating mean age does not offer significant improvement for the Asian sample used in the study. Overbury et al., 2009 (47) detect weak directional asymmetry, with the right side appearing slightly older than the left. However, the authors claim that the accuracy of the Suchey–Brooks system is not affected by this directional asymmetry. Nevertheless, they suggest selecting the “older” side as a more accurate predictor, reporting an increase in accuracy from 78% to 91% when the older side is used. Lottering et al., 2013 (17) report that the right side is more accurate when the Suchey–Brooks method is applied to Australian populations.

This study uses a suite of recently proposed computational methods that are based on coordinate data from 3D laser scans, implements three new shape-analysis algorithms, and applies five univariate and multivariate regression models for age-at-death estimation. It evaluates these methods’ sensitivity to within-individual asymmetry. These “fully computational” methods are chosen for such analysis over the gross visual assessment or bone-phase/stage matching methodologies of more familiar practice for the fact that (i) they were first proposed to eliminate subjectivity in bone evaluation and reduce error in age estimation and (ii) they are designed to capture shape information using high density, 3D coordinates and methods of continuous data analysis. Further, unlike the phase-based methods, the computational algorithms produce continuous age-at-death estimations and are thus likely to capture more subtle differences between the left and right sides.

Materials and Methods

This study uses 3D laser scan data of left and right pubic symphyses. The pubic symphysis scans were taken with the widely available desktop laser scanner, NextEngine 3D Scanner Ultra HD, following the scanning protocol from the papers of Slice and Algee-Hewitt (20) and Stoyanova et al. (21,22) to produce a high-definition 3D model: 16 divisions, 67K points per square inch with a triangle size of 0.0050, and bracket scanning

in macro mode. Figure 1 shows one of the pubic symphysis scans available in the dataset when displayed as a smooth surface and as a number of small adjacent triangles. The laser scanner generates an output file that includes a list of x , y , and z coordinates for each vertex in the 3D space. This representation of the data as a number of 3D coordinates allows for its analysis by numerical algorithms.

This study samples 176 3D laser scans of both the left and right pubic symphyses for 88 modern American males of reported White identity and with known ages-at-death. These scans were collected from the W. M. Bass Donated Skeletal Collection, housed at the University of Tennessee, Knoxville (50) and from the Maxwell Museum’s collection curated by the University of New Mexico (51). The ages in the sample range from 19 to 93 years. Figure 2 offers a histogram of the age distribution of males in the sample. In addition, height and weight are available for approximately 71 of the individuals from which body mass index (BMI) was calculated using the formula body mass (kilograms)/stature (meters)² (52). Based on the BMI, the following clinical categories were defined: underweight (BMI < 18.5), normal weight (18.5 ≤ BMI ≤ 24.99), overweight (25 ≤ BMI ≤ 29.99), and obese (BMI ≥ 30) (53). In the sample, nine individuals were classified as underweight, 27 as normal weight, 17 as overweight, and 18 as obese. The individuals’ reported height, weight, calculated BMI, and known age were used as a possible way of explaining higher asymmetry values observed in some cases.

The morphology of each pubic symphysis scan was quantified using the algorithms included in the open source, multiplatform program *forAge* (54) written specifically for the implementation of the computational, shape-based age estimation methods tested here. Given a 3D laser scan of the pubic symphysis as input, *forAge* calculates three shape scores: (i) *SAH-Score* (20), (ii) bending energy (BE) score (21) and (iii) ventral curvature (VC) score (22). The *SAH-Score* measures the variance on the surface of the pubic symphysis such that high variance values are associated with younger individuals for whom the symphyseal face is covered by furrows and ridges and low variance is associated with older individuals for whom the furrows and ridges have flattened over time. BE uses thin plate splines (TPS) to model the bending of an infinitely thin plate to match the surface of the

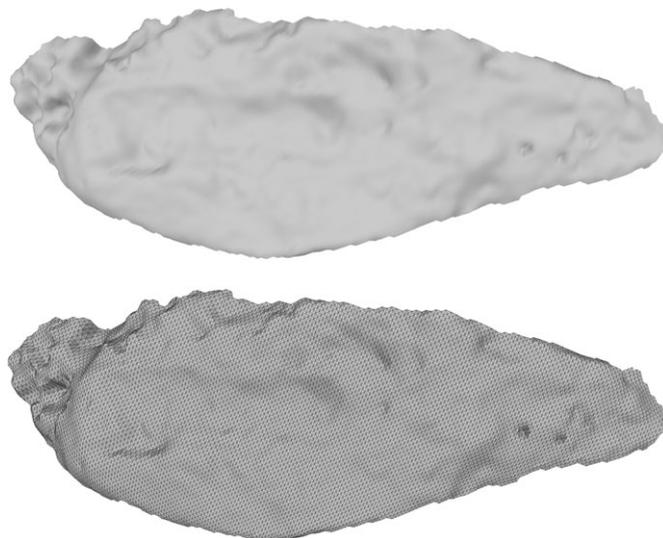


FIG. 1—An example of a 3D laser scan of the pubic symphysis shown as a smooth surface (top) and as a number of adjacent triangles (bottom).

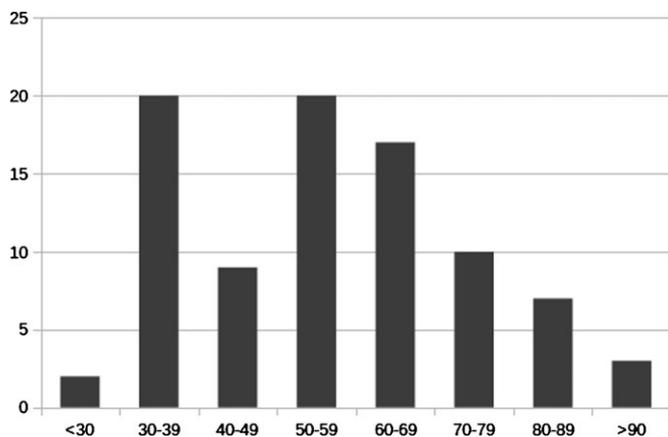


FIG. 2—A histogram of the age distribution of individuals in the data.

pubic symphysis and calculates the minimum energy required for the transformation. Similar to the *SAH-Score*, high BE values are calculated for younger individuals and lower BE values are calculated for older individuals. The VC score quantifies the outline of the ventral margin of the pubic symphysis. The ventral margin becomes more curved with aging causing low curvature values to be associated with younger people and high curvature values to be associated with older individuals. In addition to the three measures, *forAge* also calculates age-estimates using five regression models—three univariate models based on each of the three scores and two multivariate models combining the *SAH-Score* with the VC and the BE measure with the VC. All five models proposed by Stoyanova et al., 2017 (22) are built on a sample of 93 randomly selected left and right scans (one scan per individual) of White males with known ages-at-death. The sample includes approximately the same number of left and right scans. A small number (approximately 25) of the scans in the original *forAge* models were also included in this study's sample of 176 scans. The final estimates produced by all *forAge* models are point estimates of age.

In this study, the classic bias and inaccuracy measures by Saunders et al., 1992 (55) were slightly modified such that two different sets of inaccuracy and bias relevant to the evaluation of bilateral asymmetry in the pubic symphysis can be produced. The first set of inaccuracy and bias assesses the difference in raw shape-measure *values/scores* between the left side and the right side without considering age-estimates and documented ages. The second set of inaccuracy and bias evaluates the difference in *age-estimates* between the left side and the right side. After establishing the comparability of the age estimation models produced from either side, the difference between age-estimates and documented chronological ages is assessed to determine which side produces more accurate age-at-death estimates. The following paragraphs explain these steps in greater detail.

Given that the age-estimates produced by *forAge* are based on mixed left and right scans, a necessary first step of this analysis is to investigate whether the use of models based on only left or only right scans is more appropriate. This is performed by a bias and inaccuracy analysis that assesses whether one side consistently produces higher (or lower) scores and quantifies the difference between the left and right scores. For this study, the sample was split into age groups that have approximately the same number of people; specifically, the age groups were 19–39 years (22 people), 40–54 years (23 people), 55–69 years (23 people), and 70-and older (20 people). The bias is defined as

This resource was prepared by the author(s) using Federal funds provided by the U.S.

Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$$\text{Bias} = \sum (\text{left score} - \text{right score})/N,$$

and the inaccuracy as

$$\text{Inaccuracy} = \sum |\text{left score} - \text{right score}|/N,$$

where N is the number of sample points. As the three shape scores—*SAH-Score*, BE and VC—show a linear relationship to the exact age-at-death, it is implied that consistently higher (or lower) scores will result in consistent over- (under-) estimation. Similarly, a larger difference between the scores produced by the left and right side is expected to result in a larger difference between the age-estimates, hence, a larger asymmetry value.

Additionally, based on the three shape scores, new regression models were built using only left and only right scans (from the 88 individuals included in this study's sample) to assess whether the age estimation error produced by models using only one side offered any improvement compared to the mixed models. To assess whether there was a significant difference between the age-estimates from the models based on only left and only right scans, the two sets of estimates (left and right) were subjected to Student's *t*-test for paired samples, the Wilcoxon signed-rank test—as a nonparametric equivalent, the Kolmogorov–Smirnov test for distributional differences, and Spearman correlation analysis to compare the two sets of age-estimates.

The results of the analysis (presented in detail in the next section) indicate that the mixed models implemented in *forAge* are adequate. Therefore, the five age-estimates for each left and right scan were calculated using the *forAge* program for a total of 880 age-estimates. The difference between the exact age-at-death and the left estimate was compared to the difference between the exact age-at-death and the right estimate to assess whether one side consistently produced lower estimation errors. In addition, the left and right age-estimates were subjected to a bias and inaccuracy analysis. Bias is defined as

$$\text{Bias} = \sum (\text{left estimate} - \text{right estimate})/N,$$

and the inaccuracy as

$$\text{Inaccuracy} = \sum |\text{left estimate} - \text{right estimate}|/N,$$

where N is the number of individuals in the sample.

Similar to Overbury et al., 2009 (47), the number of more accurate “older” estimates was compared to the number of more accurate “younger” estimates to test whether the “older” or “younger” side offered consistently better age approximation.

Next, the difference in years between the left and right age-estimates was calculated and used as a measure of the within-individual asymmetry. Lastly, the asymmetry measure (the calculated difference in years between the left and right age-estimates) was regressed against the individuals' weights, heights, BMIs, and true ages to investigate any relationship that the asymmetry has with these life history factors.

Results

Bias and Inaccuracy Analysis of the Three Shape Scores

The results from the bias and inaccuracy analysis for the three shape scores are reported in Table 1. Overall, the inaccuracy

TABLE 1—Bias and inaccuracy between the left and right sides for each measure are reported for the entire sample and by age interval.

Age Interval and Count	SAH-Score		Bending Energy		Ventral Curvature	
	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy
Entire sample (88)	0.0005	0.0311	0.0185	0.5950	0.0022	0.0195
19–39 (21)	0.0244	0.0645	0.0124	0.4992	–0.0016	0.0164
40–54 (23)	–0.0110	0.0657	–0.2044	0.5482	0.0004	0.0211
55–69 (23)	–0.0081	0.0729	0.3579	0.6363	0.0082	0.0212
70–93 (20)	–0.0025	0.0792	–0.1088	0.7070	0.0015	0.0192
>40 (66)	–0.0074	0.0723	0.0205	0.6270	0.0034	0.0205

Next to the age interval (in parenthesis) is the count of individuals in the given sub-sample.

slightly increases with age for the *SAH-Score* and the BE. For the VC score, it increases up to 70-years and then there is a slight decrease. Figure 3 shows the absolute value of the difference between the left and right scores per individual for each of the three measures. The differences are plotted versus the individuals' ages. For ease of reading the data, the different age cohorts are marked with a different symbol. The bias values for the entire sample are all positive. This implies that the scores produced by the left sides are overall higher. However, there is a fluctuation between positive and negative values in the different age intervals, which indicates that the left sides do not consistently produce higher scores.

Assessing Whether One Side Better Models the Age-Related Morphologies

The regression models built using only left and only right surfaces produce five sets of paired (left and right) age-estimates. The paired t-test for mean difference, Kolmogorov-Smirnov test for distributional difference, Wilcoxon signed-rank test for median difference, and Spearman correlation analysis indicate that there is no significant difference between the two sets (left and right) of age-estimates. More specifically, the tests' *p*-values range from 0.1 to 0.98 and the correlation coefficients range from 0.2 to 0.74, implying that neither side offers a better model of the age-progressive changes. This justifies the mixed models offered by *forAge* on which the rest of the presented results are based.

Assessing Whether the Left or Right Age-estimate is Closer to the True Age-at-Death

The difference between the left estimate (calculated by *forAge*) and the exact age is calculated and compared to the difference between the right estimate and the exact age-at-death. Results of the number of individuals for whom the left or right estimate is more accurate (closer to the exact age-at-death) are presented in Table 2. For two of the models, univariate *SAH-Score* and multivariate *SAH-Score* and VC, there are more left sides that are closer to the true age. For the remaining three models, more of the right sides appear to be closer to the true age. However, in all cases, the difference between the number of more accurate left and right scans is less than 7% with the exception of the *SAH-Score* combined with the VC. In this latter case, the difference is about 11%. One side does not offer consistently better estimates than the other.

The majority of ages are underestimated by the computational algorithms, and therefore, the older estimate (whether left or right) is closer to the true age-at-death. In the cases, when the

models do not underestimate the age (only 165 out of 880 age-estimates), there is a fluctuation between the older and the younger estimate being closer to the true age.

Comparison of the Two Sets of Age-estimates—Left and Right

Table 3 presents a breakdown of the age differences between the left and right estimates for the 88 individuals. The majority of differences between the left and right scans are within 5 years and as few as 1% have over 15 years difference. There seems to be a slightly higher level of asymmetry for the age-estimates when using the *SAH-Score* alone and in combination with the VC measure. However, overall, the age-estimates of the left and right scans are close in value.

The results of the bias and inaccuracy analysis of the left and right age-estimates are presented in Table 4. Similar to the analysis of bias and inaccuracy in the calculated scores, there is a fluctuation between positive and negative bias values. This finding indicates that neither side (left or right) produces consistently younger or older age-estimates. The bias values for all models for postmiddle aged people are positive, implying that the left age-estimate is more often older for this cohort. However, when considering subsamples in that same age group, there is still fluctuation. The inaccuracy values are consistent among all age groups, which suggests that the age-estimates produced by the three shape scores are not sensitive to the within-individual asymmetry.

Determining Whether the Asymmetry is Related to Weight, Height, BMI, or Advanced Age

The age differences between the left and right estimates for each of the five models are calculated and regressed against the weight, height, BMI, and true ages of the individuals. The statistics show no relationship between the differences and any of the four factors. More specifically, the *p*-values of all regression models range from 0.21 to 0.94.

As the population in the dataset can be categorized more easily based on their age and BMI, the study reports these results in more details. Table 5 reports in greater detail the relationship between the true ages of the individuals and the within-individual asymmetry. Based on their age, the sample is categorized as premiddle aged (younger than 40 years) and postmiddle aged (older than 40) people. The reports show how many individuals younger than 40 have an asymmetry up to 5 years, from 5 to 10 years, from 10 to 15 years, and over 15 years for each of the five models. The same results are reported for the males who are over 40 years old. There are 22 premiddle aged and 66 postmiddle aged individuals. The findings indicate that

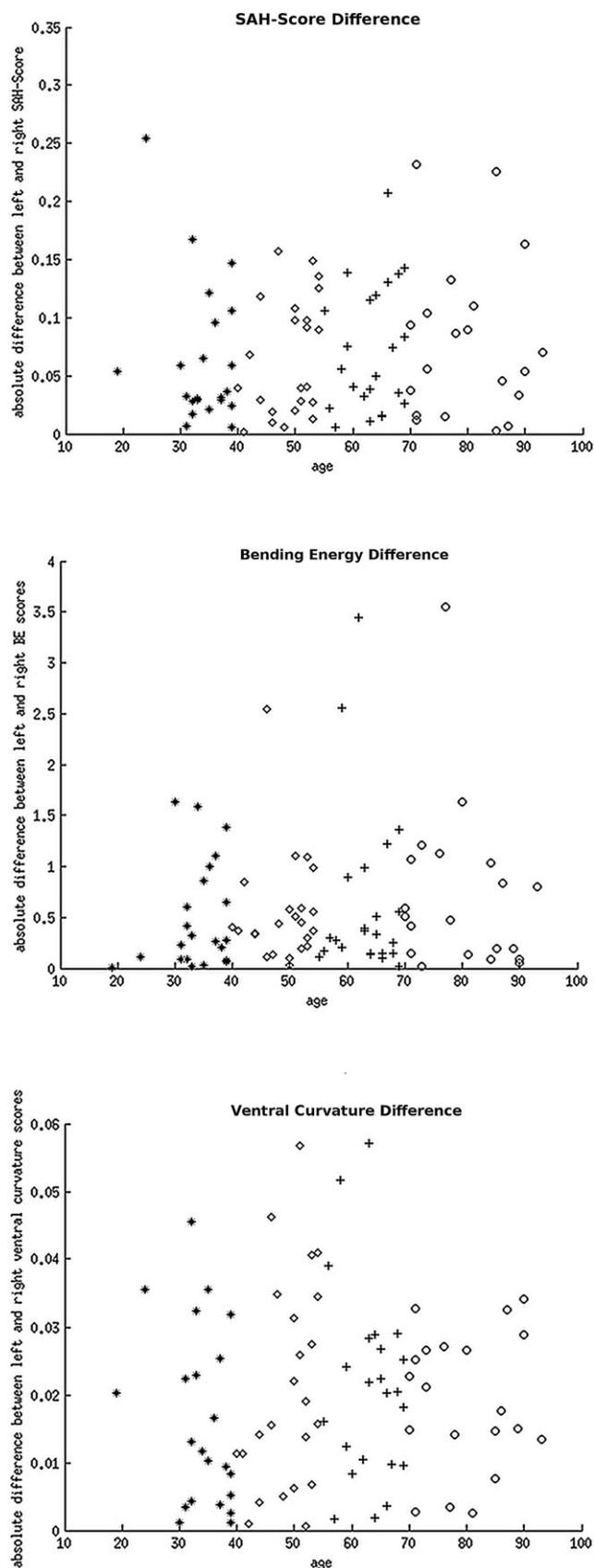


FIG. 3—The absolute value of the difference between the left and right scores for each measure plotted against the known ages-at-death. The individuals who are younger than 40 years are displayed as stars (*), those who are between 40 and 55 years are marked as diamonds, 55–70 years are marked as plus signs, and those older than 70 years are displayed as circles.

asymmetry is not affected by advanced age. For both groups (younger and older than 40), up to 1.5% of the differences are estimated to be over 15 years with the exception of the *SAH-Score* age-estimate and the age-estimate from the multivariate model including the *SAH-Score* and the VC. Even though these two models result in more people (about 10%) having an asymmetry of over 15 years, the findings are the same for both age groups.

The same breakdown of age differences—<5 years, 5–10 years, 10–15 years, or 15 years—for each of the five models is reported for each BMI category (underweight, normal weight, overweight, and obese) in Table 6. There is no notable correlation between the individual asymmetry and the BMI category.

Discussion

Age does not account for all the specific morphology of the pubic symphysis. It is a known fact that an individual's lifestyle, childbirth, and disease among other factors affect the rate at which the pubic symphysis ages (56–59). In most situations, whether forensic or bioarchaeological in nature, prior information on the individual is unknown to the observers. Therefore, it is crucial that practitioners are provided with tools for age-at-death estimation that are independent of the potentially unknown life history of the person in question. Further, in many cases, when only fragmentary remains are available, it is important to know whether both symphyseal surfaces can be used with the same level of confidence.

In this study, bias and inaccuracy analysis finds that neither side produces consistently higher/lower shape scores than the other. This suggests that both sides model the age-progressive changes of the morphology of the pubic symphysis equally well. The models using only left or only right pubic symphysis scans show similar relationship between the scores and the true age-at-death. Therefore, it can be concluded that the proposed models built on randomly selected left or right scans are sufficiently reliable, regardless of the chosen side. When the ages-at-death are estimated using the left and right side for the same individual, it is observed that the differences between the estimates produced by the two sides for the majority of individuals are within 5 years (46–76%), 75–95% are within 10 years, and 89–99% are within 15 years. Only 1–11% are more than 15 years apart. Overall, the age-estimates produced by the BE measure and its combination with the VC show to be least influenced by the within-individual asymmetry. In this context, models based on these variables should be preferred even though the *SAH-Score* generally produces a lower root-mean-square error (22). It should be noted that the computational methods may be assessing different bone features from the phase-based systems, and this could be the reason for their low-dependence on asymmetry. Further, it is important to recall that *forAge* produces point estimates of age and not age ranges and the asymmetry analyses are based on these single values.

For those individuals with notably more asymmetric sides, this study investigates whether the asymmetry is related to factors recognized as accelerating aging (47,60,61) such as the person's weight, height, advanced age, and calculated BMI. The results indicate that the difference between the age-estimates produced by the computational algorithms is not associated with any of the four factors, which may in part be due to the fact that the weight and heights are self-reported. Further, an individual's weight and height at death may not represent those variables

TABLE 2—Number of more accurate left sides versus more accurate right sides. Accuracy is defined as the difference between the known and the estimated age-at-death.

More Accurate Side	SAH-Score		TPS/BE		Ventral Curvature		SAH-Score + Outline		BE + Outline	
	Count	%	Count	%	Count	%	Count	%	Count	%
Left	47	53.4	43	48.9	42	47.7	49	55.7	41	46.6
Right	41	46.6	45	51.1	46	51.3	39	44.3	47	53.4

TABLE 3—Number of people (as a count and as a percentage) for whom the estimated age difference between the left and the right pubic symphysis is within 5, 6–10, 11–15, or >15 years.

Difference in Years	SAH-Score		TPS/BE		Ventral Curvature		SAH-Score + Outline		BE + Outline	
	Count	%	Count	%	Count	%	Count	%	Count	%
≤5	40	45.5	67	76.1	49	55.7	43	48.9	64	72.7
6–10	31	35.2	17	19.3	30	34.1	23	26.1	17	19.3
11–15	9	10.2	3	3.4	7	8	12	13.6	6	6.8
>15	8	9.1	1	1.2	2	2.2	10	11.4	1	1.2

TABLE 4—Bias and inaccuracy between the left and right estimates for each model are reported for the entire sample and by age interval.

Age Interval and Count	SAH-Score		Bending Energy		Ventral Curvature		SAH + VC		BE + VC	
	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy
Entire sample (88)	−0.2443	6.9582	0.0140	3.5517	0.6342	5.0785	−0.0718	7.0722	0.4724	3.9331
19–39 (22)	−4.9459	8.2904	−0.5668	3.0831	−0.1813	4.3686	−4.3977	7.4622	−0.7072	2.8136
40–54 (23)	2.08621	5.2891	1.5895	3.5956	0.0991	5.7121	1.2347	6.7330	2.0082	4.0717
55–69 (23)	1.4186	7.0230	−1.7343	3.6847	2.3395	5.5065	2.1921	6.9252	−0.2378	5.0204
70–93 (20)	0.3625	7.3375	0.8515	3.8635	0.1855	4.6385	0.5805	7.2025	0.8205	3.7545
>40 (66)	1.3228	6.5140	0.2075	3.7078	0.9060	5.3151	1.3701	6.9422	0.8656	4.3062

Next to the age interval (in parenthesis) is the count of individuals in the given sub-sample.

TABLE 5—Number of people who are younger or older than 40 years for whom the estimated age difference between the left and the right pubic symphysis is within 5, 6–10, 11–15, or >15 years.

Individuals Age	Difference in Years	SAH-Score		TPS/BE		Ventral Curvature		SAH-Score + Outline		BE + Outline	
		Count	%	Count	%	Count	%	Count	%	Count	%
Younger than 40	≤5	6	27.3	18	81.8	14	63.6	9	40.9	18	81.8
	6–10	9	40.9	3	13.6	6	27.3	6	27.3	4	18.2
	11–15	5	22.7	1	4.5	2	9.1	5	22.7	0	0
	>15	2	9.1	0	0	0	0	2	9.1	0	0
Older than 40	≤5	34	51.5	49	74.2	35	53	34	51.5	46	69.7
	6–10	22	33.3	14	21.2	24	36.4	17	25.8	13	19.7
	11–15	4	6.1	2	3.1	5	7.6	7	10.6	6	9.1
	>15	6	9.1	1	1.5	2	3	8	12.1	1	1.5

The results are shown as a count and as a percentage of the whole sample.

during a person's lifetime. This study's results are in disagreement with the conclusions of Wescott (61), whose study investigated the effects of obesity on ages estimated with Suchey–Brooks methods for two BMI-based cohorts of modern Americans. While they report how the obese group shows more bias, less precision, and weaker correlations between inferred and true age, the results for the pubic symphysis tests were not statistically significant. The results also disagree with those reported by Merritt (60), who found that body size was an important factor for the validity of the estimates produced using the Suchey–Brooks method. Specifically, “short and light” persons were

consistently underaged, while “tall and heavy” persons were consistently overaged. The contradictory findings reported in this study are important when seeking to infer skeletal age-at-death in the human identification context as they offer the forensic anthropologist a valuable advantage. We show how robust estimates can be produced even when life history information is unknown for the forensic case and cannot be accounted for statistically in the estimation model.

The present study agrees with the conclusions in Schmitt (48), who determined that taking the mean of the two age-estimates (left and right) does not offer an improvement to the results.

TABLE 6—A breakdown of the age difference between the left and right age-estimates for the four body mass index categories for each of the five estimation models.

Age-estimate Model	Difference in Years	Underweight	Normal Weight	Overweight	Obese
SAH-Score	≤5	3	17	7	8
	6–10	2	5	7	6
	11–15	1	2	3	2
	>15	3	3	0	2
Bending energy	≤5	7	20	13	14
	6–10	1	5	4	4
	11–15	1	2	0	0
	>15	0	0	0	0
Ventral curvature	≤5	6	10	10	12
	6–10	3	13	5	5
	11–15	0	3	1	1
	>15	0	1	1	0
SAH-Score and ventral curvature	≤5	4	14	8	10
	6–10	1	9	4	3
	11–15	1	1	5	2
	>15	3	3	0	3
Bending energy and ventral curvature	≤5	7	19	11	16
	6–10	2	4	5	1
	11–15	0	4	1	1
	>15	0	0	0	0

Using the mean age as an estimate would work in situations when one side overestimates the true age while the other side underestimates it. Here, only 33 of the total number of 880 age-estimates appear to underestimate the age while the rest to overestimate. In other words, improvements offered by the mean age can only be expected for <4% of the estimates. This study is somewhat in line with the findings of Overbury et al. (47), that proposed the use of the older side as a more accurate predictor. The reason for this agreement is that the majority of the ages estimated by the computational methods in the present study underestimate the true age. Moreover, it is shown here that selecting the older side does not necessarily offer a significant improvement to the accuracy of the methods as the left and right age-estimates are overall close in value for most individuals in the sample.

The results of the current study disagree with previous findings that have shown high prevalence of asymmetry for the Suchey–Brooks system as noted in over 60% of the samples analyzed (17,47–49). Specifically, this study does not detect either (i) the greater magnitude of asymmetry with increased age or (ii) the directional asymmetry favoring the right side (17,47). This study is, however, in line with the recent research by Kurki (62) that demonstrated low directional and mosaic patterns of asymmetry in pelvic measurements. Although any direct comparison of studies using different variables of interest (i.e., metric dimensions vs. morphological patterns of the pelvic element) should be made with caution, a few biomechanical theories can be extended to explain the trends observed in the current study. It is already well-established that different skeletal elements and bone features exhibit different degrees of asymmetry (63). Particularly, it is known that biomechanical stress, one of the causal factors for skeletal asymmetry, has greater impact on bone diaphysis rather than epiphyses (64–67), suggesting there is less environmental plasticity in the epiphyseal/articulating region and that variation in this region is more genetically constrained (68,69). In line with these findings, the lack of both directional and absolute asymmetry in age-estimates observed in the paired pubic symphyses in our study is possibly due to the bone's resilience to biomechanical stress/physical loading, a conclusion

supported by the low correlations between individual body size and age-at-death observed for the sample.

While this study provides new information on the presence of asymmetry in the pubic symphysis and its effect on age estimation when adopting a fully computational approach, it represents a first stage analysis. There are considerable opportunities for expansion, such that future work should investigate the replicability of these results and/or produce additional findings when, for example, enlarging the reference database included in the *forAge* software, using more target samples of mixed sex and ancestry, and considering age ranges instead of point estimates of age-at-death. Further, as for any study that seeks to provide recommendations for forensic anthropological practice, it is important to pursue downstream analyses that perform similar tests in an applied casework setting.

Acknowledgements

We thank Dr. Dawnie Steadman at the University of Tennessee and Dr. Heather Edgar at University of New Mexico for granting access to the skeletal materials and scanning resources that facilitated this work. We also express our deepest appreciation to the donors and their families who have contributed to the W. M. Bass and Maxwell Skeletal Collections—without their appreciation for the advancement of science through skeletal research, studies such as this would not be possible.

References

- Algee-Hewitt BFB. Age estimation in modern forensic anthropology. In: Tersigni-Tarrant MA, Shirley NR, editors. *Forensic anthropology: a comprehensive introduction*. Boca Raton, FL: CRC Press, 2017;381–20.
- Steadman DW, Adams BJ, Konigsberg LW. Statistical basis for positive identification in forensic anthropology. *Am J Phys Anthropol* 2006;131(1):15–26.
- Ousley S. Should we estimate biological or forensic stature? *J Forensic Sci* 1995;40(5):768–773.
- Milner G, Boldsen J. Skeletal age estimation: where we are and where we should go. In: Dirkmaat DC, editor. *A companion to forensic anthropology*. West Sussex, U.K.: Blackwell Publishing, 2012;224–238.
- Milner GR, Boldsen JL. Estimating age and sex from the skeleton, a paleopathological perspective. In: Grauer AL, editor. *A companion to paleopathology*. West Sussex, U.K.: John Wiley & Sons Ltd, 2012;268–284.
- Baccino E, Schmitt A. Determination of adult age at death in the forensic context. In: Schmitt A, Cunha E, Pinheiro J, editors. *Forensic anthropology and medicine: complimentary sciences from recovery to cause of death*. Totowa, NJ: Humana Press, 2006;259–280.
- Buckberry J. The (mis) use of adult age-estimates in osteology. *Ann Hum Biol* 2015;42(4):323–331.
- Cunha E, Baccino E, Marrille L, Ramsthaler F, Prieto J, Schuliar Y, et al. The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 2009;193(1–3):1–13.
- Franklin D. Forensic age estimation in human skeletal remains: current concepts and future directions. *Leg Med* 2010;12(1):1–7.
- Garvin HM, Passalacqua NV. Current practices by forensic anthropologists in adult skeletal age estimation. *J Forensic Sci* 2012;57(2):427–433.
- Márquez-Grant N. An overview of age estimation in forensic anthropology: perspectives and practical considerations. *Ann Hum Biol* 2015;42(4):308–322.
- Boldsen JL, Milner GR, Konigsberg LW, Wood JW. Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge, U.K.: Cambridge University Press, 2002;73–106.
- Buckberry JL, Chamberlain AT. Age estimation from the auricular surface of the ilium: a revised method. *Am J Phys Anthropol* 2002;119(3):231–239.
- DiGangi EA, Bethard JD, Kimmerle EH, Konigsberg LW. A new method for estimating age-at-death from the first rib. *Am J Phys Anthropol* 2009;138(2):164–176.

15. Konigsberg LW. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann Hum Biol* 2015;42(4):368–378.
16. Konigsberg LW, Frankenberg SR, Liversidge HM. Optimal trait scoring for age estimation. *Am J Phys Anthropol* 2016;159(4):557–576.
17. Lottering N, Macgregor DM, Meredith M, Alston CL, Gregory LS. 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* 2013;150(3):386–399.
18. Milner GR, Boldsen JL. Transition analysis: a validation study with known-age modern American skeletons. *Am J Phys Anthropol* 2012;148(1):98–110.
19. San-Millán M, Rissech C, Turbón D. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *Int J Legal Med* 2017;131(2):501–525.
20. Slice DE, Algee-Hewitt BF. Modeling bone surface morphology: a fully quantitative method for age-at-death estimation using the pubic symphysis. *J Forensic Sci* 2015;60(4):835–843.
21. Stoyanova D, Algee-Hewitt BF, Slice DE. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am J Phys Anthropol* 2015;158(3):431–440.
22. Stoyanova DK, Algee-Hewitt BF, Kim J, Slice DE. 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* 2017;62(6):1434–1444.
23. Villa C, Buckberry J, Lynnerup N. Evaluating osteological ageing from digital data. *J Anat*. <https://doi.org/10.1111/joa.12544>. Epub 2016 Sep 13.
24. Katz D, Suchey JM. Age determination of the male os pubis. *Am J Phys Anthropol* 1986;69(4):427–435.
25. Brooks S, Suchey JM. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum Evol* 1990;5(3):227–238.
26. Brooks ST. Skeletal age at death: the reliability of cranial and pubic age indicators. *Am J Phys Anthropol* 1955;13(4):567–597.
27. McKern TW, Stewart TD. Skeletal age changes in young American males: analyzed from the standpoint of age identification. *Quartermaster Research and Development Command; 764 Technical Report EP-45*. Natick, MA: U. S. Army Quartermaster Research and Development Center, 1957.
28. McKern TW. Estimation of skeletal age from combined maturational activity. *Am J Phys Anthropol* 1957;15(3):399–408.
29. Gilbert BM, McKern TW. A method for aging the female Os pubis. *Am J Phys Anthropol* 1973;38(1):31–38.
30. Suchey JM. Problems in the aging of females using the Os pubis. *Am J Phys Anthropol* 1979;51(3):467–470.
31. Todd TW. Age changes in the pubic bone. I. The male white pubis. *Am J Phys Anthropol* 1920;3(3):285–334.
32. Todd TW. Age changes in the pubic bone. *Am J Phys Anthropol* 1921;4(1):1–70.
33. Berg GE. Pubic bone age estimation in adult women. *J Forensic Sci* 2008;53(3):569–577.
34. Djurić M, Džonić D, Nikolić S, Popović D, Marinković J. Evaluation of the Suchey-Brooks method for aging skeletons in the Balkans. *J Forensic Sci* 2007;52(1):21–23.
35. Fleischman JM. A comparative assessment of the Chen et al. and Suchey-Brooks pubic aging methods on a North American sample. *J Forensic Sci* 2013;58(2):311–323.
36. Godde K, Hens SM. Age-at-death estimation in an Italian historical sample: a test of the Suchey-Brooks and transition analysis methods. *Am J Phys Anthropol* 2012;149(2):259–265.
37. Hartnett KM. 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* 2010;55(5):1152–1156.
38. Hoppa RD. Population variation in osteological aging criteria: an example from the pubic symphysis. *Am J Phys Anthropol* 2000;111(2):185–191.
39. Kimmerle EH, Konigsberg LW, Jantz RL, Baraybar JP. Analysis of age-at-death estimation through the use of pubic symphyseal data. *J Forensic Sci* 2008;53(3):558–568.
40. Kimmerle EH, Prince DA, Berg GE. Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth. *J Forensic Sci* 2008;53(3):594–600.
41. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 2008;53(3):541–557.
42. Merritt CE. A test of Hartnett's revisions to the pubic symphysis and fourth rib methods on a modern sample. *J Forensic Sci* 2014;59(3):703–711.
43. Miranker M. A comparison of different age estimation methods of the adult pelvis. *J Forensic Sci* 2016;61(5):1173–1179.
44. Sakaue K. Application of the Suchey-Brooks system of pubic age estimation to recent Japanese skeletal material. *Anthropol Sci* 2006;114(1):59–64.
45. San Millán M, Rissech C, Turbón D. A test of Suchey-Brooks (pubic symphysis) and Buckberry-Chamberlain (auricular surface) methods on an identified Spanish sample: paleodemographic implications. *J Archaeol Sci* 2013;40(4):1743–1751.
46. Savall F, Rérolle C, Hérin F, Dédouit F, Rougé D, Telmon N, et al. Reliability of the Suchey-Brooks method for a French contemporary population. *Forensic Sci Int* 2016;266:586. e1–e5.
47. Overbury RS, Cabo LL, Dirkmaat DC, Symes SA. Asymmetry of the os pubis: implications for the Suchey-Brooks method. *Am J Phys Anthropol* 2009;139(2):261–268.
48. Schmitt A. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. *Int J Osteoarchaeol* 2004;14(1):1–6.
49. Sinha A, Gupta V. A study on estimation of age from pubic symphysis. *Forensic Sci Int* 1995;75(1):73–78.
50. Bass Collection, Forensic Anthropology Center and Dept of Anthropology, University of Tennessee at Knoxville; <http://fac.utk.edu/> (accessed March 7, 2018).
51. Maxwell Museum's Documented Skeletal Collection; http://www.unm.edu/u/~osteolab/coll_doc.html (accessed March 7, 2018).
52. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *J Clin Epidemiol* 1972;25(6):329–343.
53. Centers for Disease Control and Prevention (CDC). Healthy weight. About adult BMI; https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html (accessed June 21, 2018).
54. ForAge; <http://morphlab.sc.fsu.edu/software/forAge/> (accessed March 7, 2018).
55. Saunders SR, Fitzgerald C, Rogers T, Dudar C, McKillop H. A test of several methods of skeletal age estimation using a documented archaeological sample. *Can Soc Forensic Sci J* 1992;25(2):97–118.
56. Bongiovanni R. Effects of parturition on pelvic age indicators. *J Forensic Sci* 2016;61(4):1034–1040.
57. Mays S. The effect of factors other than age upon skeletal age indicators in the adult. *Ann Hum Biol* 2015;42(4):332–341.
58. Campanacho V, Santos AL, Cardoso HF. Assessing the influence of occupational and physical activity on the rate of degenerative change of the pubic symphysis in Portuguese males from the 19th to 20th century. *Am J Phys Anthropol* 2012;148(3):371–378.
59. Cunningham PM, Brennan D, O'Connell M, MacMahon P, O'Neill P, Eustace S. Patterns of bone and soft-tissue injury at the symphysis pubis in soccer players: observations at MRI. *Am J Roentgenol* 2007;188(3):W291–W296.
60. Merritt CE. The influence of body size on adult skeletal age estimation methods. *Am J Phys Anthropol* 2015;156(1):35–57.
61. Wescott DJ, Drew JL. Effect of obesity on the reliability of age-at-death indicators of the pelvis. *Am J Phys Anthropol* 2015;156(4):595–605.
62. Kurki HK. Bilateral asymmetry in the human pelvis. *Anat Rec* 2017;300(4):653–665.
63. Auerbach BM, Ruff CB. Limb bone bilateral asymmetry: variability and commonality among modern humans. *J Hum Evol* 2006;50(2):203–218.
64. Churchill SE, Formicola V. A case of marked bilateral asymmetry in the upper limbs of an upper palaeolithic male from Barma Grande (Liguria), Italy. *Int J Osteoarchaeol* 1997;7(1):18–38.
65. Trinkaus E, Churchill SE, Ruff CB. Postcranial robusticity in Homo. II: Humeral bilateral asymmetry and bone plasticity. *Am J Phys Anthropol* 1994;93(1):1–34.
66. Ruff CB, Jones HH. Bilateral asymmetry in cortical bone of the humerus and tibia—sex and age factors. *Hum Biol* 1981;53(1):69–86.
67. Sakaue K. Bilateral asymmetry of the humerus in Jomon people and modern Japanese. *Anthropol Sci* 1997;105(4):231–246.
68. Lieberman DE, Devlin MJ, Pearson OM. Articular area responses to mechanical loading: effects of exercise, age, and skeletal location. *Am J Phys Anthropol* 2001;116(4):266–277.
69. Lanyon L. The influence of function on the development of bone curvature. An experimental study on the rat tibia. *J Zool* 1980;192(4):457–466.

11-16-2018

Thinking Computationally about Forensics: Anthropological Perspectives on Advancements in Technologies, Data, and Algorithms

Bridget F.B. Algee-Hewitt
Stanford University, bridgeta@stanford.edu

Jieun Kim
Florida State University, eun3022@gmail.com

Cris E. Hughes
University of Illinois at Urbana-Champaign, hughesc@uiuc.edu

Recommended Citation

Algee-Hewitt, Bridget F.B.; Kim, Jieun; and Hughes, Cris E., "Thinking Computationally about Forensics: Anthropological Perspectives on Advancements in Technologies, Data, and Algorithms" (2018). *Human Biology Open Access Pre-Prints*. 133. https://digitalcommons.wayne.edu/humbiol_preprints/133

This Open Access Preprint is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState.

Introduction: Thinking Computationally about Forensics: Anthropological Perspectives on Advancements in Technologies, Data and Algorithms

Bridget F. B. Algee-Hewitt,¹ Jieun Kim,^{2*} and Cris E. Hughes³

¹Center for Comparative Studies in Race & Ethnicity, Stanford University, Stanford, California, USA.

²Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA.

³Department of Anthropology, University of Illinois at Urbana–Champaign, Champaign, Illinois, USA.

*Correspondence to: Jieun Kim, Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306-4120 USA. Email: jkim17@fsu.edu.

Short Title: Thinking Computationally about Forensics

KEY WORDS: FORENSIC ANTHROPOLOGY, QUANTITATIVE METHODS, MEDICO-LEGAL CASEWORK, HUMAN IDENTIFICATION, SKELETAL ANALYSIS, GENETIC ANALYSIS, REMAINS RECOVERY, DIVERSE POPULATIONS.

While forensic anthropology is often characterized as an applied science, it is deeply rooted in the larger discipline of biological anthropology. As forensic practitioners, we work to extend the theory of, and methods for, the study of human variation to the medico-legal context. We continue, therefore, to address the fundamental questions of topics critical to biological anthropology, such as the degree and distribution of skeletal and genetic diversity and the effects of environment and life history on morphological expression, just as we seek to infer the demographic parameters of sex, age, and ancestry that allow us to broadly characterize modern peoples. In speaking for the single person, however, forensic anthropologists are uniquely challenged with the issue of scale. We must distill down our approaches (or methods) of biological anthropology for the detection and documentation of populational trends to the level of the individual forensic case as we reconstruct the biological profile and address the personal identity concerns that dominate the forensic anthropological analysis of unknown human remains. In concert with this change in scope, forensic anthropologists must also contend with a refocusing of perspective towards the investigative and judicial system. At once, we are expected to respond to the dynamic needs of individual identification, in the service of human rights, social justice, and the medico-legal community, the increasing demands for scientific rigor in case analysis and reporting, and the changing expectations for admissible evidence and expert testimony in the courtroom (Christensen and Crowder 2009; Grivas and Komar 2008; Lesciotta 2015; Steadman 2009; Steadman et al. 2006; Wiersema et al. 2009).

Not surprisingly, the field of forensic anthropology has evolved considerably over the course of the last half-century. The present state of the discipline is very different from when it was first admitted into the American Academy of Forensic Sciences as the ‘physical anthropology’ section in 1972, yet there are interesting and important parallel moments in its

developmental trajectory. A decade later, Snow (1982) wrote about the new expansion of *physical anthropology into the area of forensics* “at a time when many physical anthropologists are deeply concerned with the need to expand the scope of our field beyond its traditional boundaries [...]. (p. 97).” Similarly, we are writing here about how *forensic anthropology is presently undergoing its own transformation*, as it expands its reach to adopt methods that would have been once outside the limits of the forensic anthropologist’s traditional expertise. As a field, it stands, therefore, at the crossroads between diverse areas of study beyond skeletal variation—from computational biology, genetics, isotopes, and demography, to cultural anthropology (Algee-Hewitt et al. 2018; Hughes et al. 2017). This point is most obviously supported by the amendment of its AAFS section title to the more inclusive title of ‘Anthropology’ in 2013. This change was equally born out by the trends observed in publications specific to forensic anthropology. From a Web of Science analysis of the publications of the *Journal of Forensic Science*, when filtered for Anthropology as research area and constrained to the roughly, five-year period of 2013 to 2018, we find that the top ten cited papers (Christensen et al. 2014; Crowder et al. 2013; Edgar 2013; Hackman and Black 2013; Hauther et al. 2015; Hefner and Ousley 2014; Kim et al. 2013; Stephan 2014; Stephan et al. 2013; Tise et al. 2013) have titles that explicitly address the computational topics of advanced statistics, classification algorithms, and imaging methods and speak to the bio-social issues of racial and ancestral identity for peoples of multiple origins with complex population histories. Moreover, in these studies, we see more multicomponent frameworks that embrace the technological advances in complementary, computationally driven, fields, and a recognition of the potential value of computational approaches to the forensic anthropologist’s applied work, whether in aid of law enforcement or

in large scale responses to humanitarian crises, human right violations or mass disasters. This movement in forensic anthropology represents swift and major progress forward.

This shift in our research priorities is especially important as it comes at the same time as the wider forensic community is undergoing a period of self-reflection and redefinition of expectations. With this has come many critiques of the rigor of the *science* of forensics, which have acted in several subfields, like fingerprint, bloodstain and bitemark analysis, as a motivating factor for change not only in protocols but also in the perception of the value of certain techniques for casework and the scientific basis for their justification as trial evidence (Cooper 2016; Expert Working Group on Human Factors in Latent Print Analysis 2012; Page et al. 2011a; Page et al. 2011b; Saks et al. 2016; Taylor et al. 2016). We recall how in 2009 the National Research Council (NRC) of the United States (Committee on Identifying the Needs of the Forensic Sciences Community and National Research Council 2009) published a report that admonished the state of the field, writing that “[t]he bottom line is simple: [i]n a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions...”(p. 53) and how “[m]uch forensic evidence [...] is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing...”(p. 107-108). This report also charged the forensic science community to make the fundamental changes necessary to drive its practitioners towards a true transformation in thinking, asking forensic scientists to make the improvements, both “systemic and scientific,” that are essential to “...ensure the reliability of the disciplines, establish enforceable standards, and promote best practices and their consistent application” (p. xix). Its coincident release at the peak year for the number of exonerations by DNA in the United States (CNN 2013) has helped us to recognize the miscarriages of justice that

can result from the uncritical admission of unvalidated science, expert testimony or evidence (Garrett 2012). With progressive interest over time, likely owing to the widespread visibility brought to these issues by, for example, the Innocence Project, this NRC report has garnered more attention in very recent years: doubling its readership since 2014, accumulating 21,787 downloads as of April 2018, and generating an Altmetric score of 612, which serves as a measure of the amount of attention the report has received from both social media and the news (The National Academies Press 2018).

We can also look to the 2016 announcement by the Department of Justice (DOJ) on scientific certainty. Building on the momentum of the NRC report and on the recommendations of the National Commission on Forensic Science (NCFS), it formally rejects the phrase “to a reasonable degree of scientific certainty,” which has long served in expert testimony as a signifier of factuality (National Commission on Forensic Science (NCFS) 2016). This mandate acknowledges what others in the legal system as well as in forensic anthropology have already noted (Kaye 2010; Steadman et al. 2006): it is a linguistic trope plagued by subjectivity, that may be erroneously equated with certainty at the level of beyond a “reasonable” doubt and that introduces confusion in the presence of actual probabilistic evidence. At the most foundational level, therefore, all forensic sciences need the transparent quantification of uncertainty through error testing and validation studies as well as approaches that reduce error and permit the probabilistic statements on the weight of evidence – these are expectations that can only be met by computational approaches.

Thinking computationally about forensics can significantly strengthen our forensic anthropological practice, yielding more reliable and accurate findings. Computational methods, in particular, offer several advantages to the study of forensic anthropological data. Through the

analysis of large quantities of informational data, they allow us, as researchers, to perform more comprehensive or deeper investigations, effectively overcoming the limitations of cognitive ability and building stronger scientific foundations for our applied casework techniques. For example, recent advancement in various innovative data capturing technologies enable us to collect and examine more nuanced and detailed information from forensic samples and scenes (Claes et al. 2014; Park et al. 2017; Perlin et al. 2015; Stoyanova et al. 2017; van Oorschot et al. 2010; Walsh et al. 2014), and the availability of free-access software and computational algorithms, through open source coding initiatives, allows for widespread access to more complex probabilistic methods (Boldsen et al. 2002; Kim et al. 2013; Konigsberg 2015; Konigsberg et al. 2016), while significantly reducing processing time for big – high dimensional, large sample – data. Most importantly, the growing number of inter-disciplinary collaborations and the high competition in the technology market have lowered the burden of accessibility to new technologies, both in terms of expense as well as practice, making approachable those new methods that were once off limits because of their sizeable learning-curve and need for specialized knowledge or degree of expertise. The ubiquity of technology that is equally relevant to science and society, therefore, pushes us forward towards embracing the new technological developments and adopting the new methodological approaches.

By probing anthropological case data in previously unavailable ways, computational tools also give us a means to reveal latent data trends, identify and explore novel questions, and establish inferential procedures that deliver more satisfying results. Thereby, we improve the quality of our aid to the medico-legal community. Finally, when computational systems are used to represent expert knowledge, they allow us to better capture, distill and interpret complex data (Algee-Hewitt 2016; Konigsberg et al. 2016), while also improving precision and accuracy,

reducing subjectivity, and facilitating the automation of traditional procedures – many of which have been shown to suffer from human, whether, intra- or inter-observer error (Kimmerle et al. 2008; Stoyanova et al. 2017). They permit, therefore, the analysis of data in a standardized way, that promotes objectivity and reproducibility, method- and/or self-assessment and the reporting of rates of error. Incorporating computational methods into the forensic anthropological toolkit helps us to overcome the common criticism that forensic evidence and courtroom testimony, in areas exclusive of, the gold standard, DNA analysis, lack a strong scientific basis (Steadman et al. 2006) To most effectively and responsibly profit from these computational resources, we as both researchers and practitioners must engage in focused discussions of their advantages and limitations. The “hybrid-intelligence of human and machines” (Franke and Srihari 2007) requires the development of best practices for their use, as we contend with the evolving issues of software compatibility and our data management, bioethical concerns over the new kinds of information that we now have or make accessible, and the question of the dissemination of results among our peers, to our students and future colleagues in the classroom, for the medico-legal community, and to the public who we seek to serve and educate.

Now is an exciting and critical time for the future of forensic anthropology. It is ripe for the field and the practitioner to embrace the challenges that increased rigor in case evaluation techniques and heightened expectations for evidence and testimony bring. As current educators, through our dissemination of research, mentorship in the lab and teaching in the classroom, it is also time for us to redefine our educational expectations that reflect the necessary integration of field and new method and, so, meet the needs of the next generation of practitioners. We are charged, therefore, to achieve a synthesis of anthropological and computational thinking. Such interdisciplinary educational opportunities will greatly benefit the field, just as interdisciplinary

research collaborations are already advancing our ability to adopt external approaches and integrate cutting-edge technologies in order to revisit prior assumptions and shed fresh light on forensic anthropological questions (Adserias-Garriga et al. 2017; Algee-Hewitt et al. 2016; Hughes et al. 2017; Villa et al. 2016).

By turning to what Steadman (2018) has called “computational anthropology,” we can benefit at present from the many advantages that marrying technology with statistically driven approaches to forensic case analysis offers while also working to propel the theory, methods and practice of forensic anthropology forward. The papers included in the two special issues on “Thinking Computationally about Forensics” that we introduce here are the products of a symposium organized by Algee-Hewitt and Kim for the 87th annual meeting of the American Association of Physical Anthropologists in Austin, TX, in 2018. Its goal was to motivate discussions between those research-focused biological anthropologists, who engage in computational anthropology and those practitioners of forensic anthropology, whose applied work serves the medico-legal community. This symposium operated as a nexus for discussions about the interactions between theoretical contributions to forensic analysis and the practical value of these advancements in the actual casework setting. These special issues embody the spirit of this symposium, as we reflect on the ways in which advanced quantitative approaches to research in biological anthropology can be extended to questions of relevance to forensic anthropology. The contributors to these issues dive deeply into a variety of subjects, from the construction of the biological profile to investigative techniques for human remains recovery, yet they all exemplify the kind of novel work which stands at the point of intersection between theory, method and practice. These articles show the promise of improved methods for ancestry estimation from skeletal metrics and genetics alike, the future of individual identifiability from

linked genetic data, the bioethical concerns over privacy and profiling as technological advancements in predictive modeling move faster than regulatory bodies can sustain their oversight, the application of new imaging technologies for enhanced investigation, whether in the context of the facial reconstruction of human remains from CT scans or the detection of clandestine burials using unnamed aerial vehicles as remote sensing platforms for data acquisition, or the novel fully-computational shape-based estimation of age at death for understudied populations.

The commentaries that accompany each issue address the places we have been, where we stand now and where we must go in service of the field. They provide us with special insight into how computational approaches to forensics are viewed from the perspective of the expert in human identification practice and in scientific computing. Although their point of reference differs, the overarching message is remarkably similar. It is not enough for us to execute computational research in an intellectual vacuum, promising improvement to current methods in forensic casework. We must also deliver on that promise by producing tools – whether as new frameworks of analysis, standards and protocols, executable code, or software – that are conceptually and physically accessible to the wider community of forensic anthropologists. In doing so, we should facilitate, with greater ease, the adoption of the methods of “computational anthropology” by the average practitioner as well their extension by the advanced user, permit the straightforward evaluation of their statistical foundations and rates of error, and encourage their acceptance by the medico-legal system.

The purpose of these special issues is to provide a forum for us to foreground our interests in advancing computational research that has implications for the forensic anthropological sciences, and to enrich current procedures with the potential to change the course

of future human identification practice. Here, we bring together a mix of participants, who engage wide-ranging skeletal, genomic, phenotypic and meta-data analyses, to 1) introduce new algorithmic advances and methodological improvement, 2) present work on the application of computational techniques to understudied populations, novel datasets or new information types, and 3) speak to the challenges that the revolution in data technologies may pose for future scientific investigation as well as the broader social effects on issues of policy, privacy and lay interpretation. We believe that the work of the issues' contributing authors exemplifies successful attempts at making advancements in all steps of the forensic casework analysis process: from locating clandestine graves, to the molecular, chemical, and osteological analysis of the retrieved remains in the forensic laboratory, to the re-fleshing of forensic case to reveal the unique features of the person, and finally to the presentation of our evidence and the communication of our findings beyond the laboratory and courtroom into the wider discourse, here in Human Biology.

We hope that this journal's present commitment to forensic research represents the first of many, with future issues that will showcase the continued development of exciting new techniques, especially for unexplored data types and underserved populations, and support the dissemination of promising research in forensic anthropology to a larger audience.

Received 2 May 2018; accepted for publication 16 May 2018.

Literature Cited

- Adserias-Garriga, J., N. M. Quijada, M. Hernandez et al. 2017. Dynamics of the oral microbiota as a tool to estimate time since death. *Mol. Oral Microbiol.* 32:511–516.
- Algee-Hewitt, B. F. B. 2016. Population inference from contemporary American craniometrics. *Am. J. Phys. Anthropol.* 160:604–624.
- Algee-Hewitt, B. F. B., M. D. Edge, J. Kim et al. 2016. Individual identifiability predicts population identifiability in forensic microsatellite markers. *Curr. Biol.* 26:935–942.
- Algee-Hewitt, B. F. B., C. E. Hughes, and B. E. Anderson. 2018. Temporal, geographic and identification trends in craniometric estimates of ancestry for persons of Latin American origin. *Forensic Anthropology* 1:4–17.
- Boldsen, J. L., G. R. Milner, L. W. Konigsberg et al. 2002. Transition analysis: A new method for estimating age from skeletons. In *Paleodemography: Age Distributions from Skeletal Samples*, R. D. Hoppa and J. W. Vaupel, eds. Cambridge: Cambridge University Press, 73–106.
- Christensen, A. M., and C. M. Crowder. 2009. Evidentiary standards for forensic anthropology. *J. Forensic Sci.* 54:1,211–1,216.
- Christensen, A. M., C. M. Crowder, S. D. Ousley et al. 2014. Error and its meaning in forensic science. *J. Forensic Sci.* 59:123–126.
- Claes, P., D. K. Liberton, K. Daniels et al. 2014. Modeling 3D facial shape from DNA. *PLoS Genet.* 10:e1004224.
- CNN. 2013. Exonerated: Cases by the numbers. Retrieved from:
<https://www.cnn.com/2013/12/04/justice/prisoner-exonerations-facts-innocence-project/index.html>.

- Committee on Identifying the Needs of the Forensic Sciences Community, and National Research Council. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, D.C.
- Cooper, S. L. 2016. Challenges to fingerprint identification evidence: Why the courts need a new approach to finality. *Mitchell Hamline Law Review* 42:756–790.
- Crowder, C., C. W. Rainwater, and J. S. Fridie. 2013. Microscopic analysis of sharp force trauma in bone and cartilage: A validation study. *J. Forensic Sci.* 58:1,119–1,126.
- Edgar, H. J. H. 2013. Estimation of ancestry using dental morphological characteristics. *J. Forensic Sci.* 58:S3–S8.
- Expert Working Group on Human Factors in Latent Print Analysis. 2012. Latent print examination and human factors: Improving the practice through a systems approach.
- Franke, K., and S. N. Srihari. 2007. Computational forensics: Towards hybrid-intelligent crime investigation. Third International Symposium on Information Assurance and Security, 383–386.
- Garrett, B. L. 2012. *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*. Cambridge, MA: Harvard University Press.
- Grivas, C. R., and D. A. Komar. 2008. Kumho, Daubert, and the nature of scientific inquiry: Implications for forensic anthropology. *J. Forensic Sci.* 53:771–776.
- Hackman, L., and S. Black. 2013. The reliability of the Greulich and Pyle atlas when applied to a modern Scottish population. *J. Forensic Sci.* 58:114–119.
- Hauther, K. A., K. L. Cobaugh, L. M. Jantz et al. 2015. Estimating time since death from postmortem human gut microbial communities. *J. Forensic Sci.* 60:1,234–1,240.

- Hefner, J. T., and S. D. Ousley. 2014. Statistical classification methods for estimating ancestry using morphoscopic traits. *J. Forensic Sci.* 59:883–890.
- Hughes, C. E., B. F. B. Algee-Hewitt, R. Reineke et al. 2017. Temporal patterns of Mexican migrant genetic ancestry: Implications for identification. *Am. Anthropol.* 119:193–208.
- Kaye, D. H. 2010. *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard University Press.
- Kim, D. I., U. Y. Lee, S. O. Park et al. 2013. Identification using frontal sinus by three-dimensional reconstruction from computed tomography. *J. Forensic Sci.* 58:5–12.
- Kimmerle, E. H., D. A. Prince, and G. E. Berg. 2008. Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth. *J. Forensic Sci.* 53:594–600.
- Konigsberg, L. W. 2015. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann. Hum. Biol.* 42:368–378.
- Konigsberg, L. W., S. R. Frankenberg, and H. M. Liversidge. 2016. Optimal trait scoring for age estimation. *Am. J. Phys. Anthropol.* 159:557–576.
- Lesciotto, K. M. 2015. The impact of Daubert on the admissibility of forensic anthropology expert testimony. *J. Forensic Sci.* 60:549–555.
- The National Academies Press. 2018. *Stats on Strengthening Forensic Science in the United States*.
- National Commission on Forensic Science (NCFS). 2016. Testimony using the term “reasonable scientific certainty.”
- Page, M., J. Taylor, and M. Blenkin. 2011a. Forensic identification science evidence since Daubert: Part I--A quantitative analysis of the exclusion of forensic identification science evidence. *J. Forensic Sci.* 56:1,180–1,184.

- Page, M., J. Taylor, and M. Blenkin. 2011b. Forensic identification science evidence since Daubert: Part II--Judicial reasoning in decisions to exclude forensic identification evidence on grounds of reliability. *J. Forensic Sci.* 56:913–917.
- Park, C. S., H. P. Jeon, K. S. Choi et al. 2017. Application of 3D laser scanner to forensic engineering. *J. Forensic Sci.* 63:930–934.
- Perlin, M. W., J. M. Hornyak, G. Sugimoto et al. 2015. TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *J. Forensic Sci.* 60:857–868.
- Saks, M. J., T. Albright, T. L. Bohan et al. 2016. Forensic bitemark identification: Weak foundations, exaggerated claims. *J. Law Biosci.* 3:538–575.
- Snow, C. C. 1982. Forensic anthropology. *Annu. Rev. Anthropol.* 11:97–131.
- Steadman, D. W. 2009. Personal identification: Theory and applications. The case study approach. In *Hard Evidence: Case Studies in Forensic Anthropology*, 2nd ed. D. W. Steadman, ed. Upper Saddle River, NJ: Prentice Hall, 1–7.
- Steadman, D. 2018. Who needs data? I've got experience! Paper presented at the 87th Annual Meeting of American Association of Physical Anthropologists. 2018 April. Austin, TX.
- Steadman, D. W., B. J. Adams, and L. W. Konigsberg. 2006. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.* 131:15–26.
- Stephan, C. N. 2014. The application of the central limit theorem and the law of large numbers to facial soft tissue depths: T-Table robustness and trends since 2008. *J. Forensic Sci.* 59:454–462.
- Stephan, C. N., E. K. Simpson, and J. E. Byrd. 2013. Facial soft tissue depth statistics and enhanced point estimators for craniofacial identification: The debut of the shorth and the 75-shormax. *J. Forensic Sci.* 58:1,439–1,457.

- Stoyanova, D. K., B. F. B. Algee-Hewitt, J. Kim et al. 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.* 62:1,434–1,444.
- Taylor, M. C., T. L. Laber, P. E. Kish et al. 2016. The reliability of pattern classification in bloodstain pattern analysis, part 1: Bloodstain patterns on rigid non-absorbent surfaces. *J. Forensic Sci.* 61:922–927.
- Tise, M. L., M. K. Spradley, and B. E. Anderson. 2013. Postcranial sex estimation of individuals considered Hispanic. *J. Forensic Sci.* 58:S9–S14.
- van Oorschot, R. A. H., K. N. Ballantyne, and R. J. Mitchell. 2010. Forensic trace DNA: A review. *Investig. Genet.* 1:14.
- Villa, C., J. Buckberry, and N. Lynnerup. 2016. Evaluating osteological ageing from digital data. *J. Anat.*
- Walsh, S., L. Chaitanya, L. Clarisse et al. 2014. Developmental validation of the HRisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci. Int. Genet.* 9:150–161.
- Wiersema, J., J. C. Love, and L. G. Naul. 2009. The influence of the Daubert guidelines on anthropological methods of scientific identification in the medical examiner setting. In *Hard Evidence: Case Studies in Forensic Anthropology*, 2nd ed. D. W. Steadman, ed. Upper Saddle River, NJ: Prentice Hall, 80–90.

TITLE:

The Development and Use of Computational Tools in Forensic Science

AUTHOR:

Dennis E. Slice

5 AFFILIATIONS:

Department of Scientific Computing, Florida State University, Tallahassee, FL, USA

Department of Anthropology, University of Vienna, Vienna, Austria

KEY WORDS:

computational forensics, morphometrics, statistical models

10 RUNNING TITLE:

Computational Tools in Forensic Science

Slice: Computational Tools in Forensic Science

ABSTRACT: Modern computational resources make available a rich toolkit of statistical methods that can be applied to forensic questions. This toolkit is built on the foundation of statistical developments dating back to the 19th century. To fully and effectively exploit these
15 developments, both the makers and users of software must be keenly aware of the quality, i.e., the accuracy and precision, of the data being modeled or analyzed, and end-users must be sufficiently familiar with the underlying theory to understand the process and results of any analysis or software they use. This is especially important for medico-legal personnel who might be called upon to testify in a court of law and be subject to cross-examination. With respect to
20 the development of computational tools, it is increasingly important that they be made available as open-source code to avoid the pitfalls of commercial software support and the potential dependence of end-users on orphaned software.

Slice: Computational Tools in Forensic Science

What might be called the classical approach to natural history, including general taxonomy, physical anthropology, and forensics, can be quite effective. This involves the proclamations and pronouncements of a seasoned professional who has carefully inspected thousands of individuals and comparative samples. A problem with this approach is that it is neither easily extensible nor scalable nor transferable to practitioners of lesser expertise. Statistical analysis, arising from probability theory and greatly elaborated in the late 19th and early-to-mid 20th centuries by the likes of Galton, Pearson, Gosset, and Fisher did much to remove the potential subjectivity of the classical process for simple analyses and measurements. The work of Pearson (1928) on the skull of Henry Stewart, Lord Darnley, through the comparison of portraits and contemporary descriptions of the same and of Pearson and Morant (1934) on the authenticity of the skull of Oliver Cromwell represent early, quantitative analyses addressing the very forensic problem of the individual identity of contested human remains. These studies outline and implement much of the morphometric methodology still useful in the toolkit of the forensic anthropologist.

Many of the interesting problems in shape analysis, including forensic applications, are, however, multivariate in nature and not readily suited to the inspection or analysis of individual measures or indices. While Galton, Pearson and others were concerned with correlation amongst variables, it required the relatively recent advent of matrix algebra and multivariate analysis to fully elucidate the principles and potential of the joint analysis of multiple measurements. Cayley defined matrix multiplication and inverses in the mid-19th century (Cayley 1858), which was exploited and extended by Wishart in the study of covariances (Wishart 1928) and used by Hotelling (1931) and Wilks (1946) to extend the univariate means testing to suites of covarying

Slice: Computational Tools in Forensic Science

variables. As with the much of the work published in *Biometrika* contributing to the
45 development of univariate statistics, physical anthropology played a key role as, for instance,
Mahalanobis' work extending Pearson's Coefficient of Racial Likeness to account for covariance
structure and leading to the now ubiquitous Mahalanobis distance (1927, 1936). In the realm of
modern shape analysis, Mosier (1939) provided a method for comparing configurations of points
in any dimension that is the fundamental operation used in the Generalized Procrustes Analysis
50 of shape (Gower 1975, Rohlf and Slice 1990).

Despite the power and desirability of the approaches developed and available in the early 20th
century, their sophistication and computational requirements limited their general adoption and
application to practical problems of interest to physical and forensic anthropologists – Howells'
(1973) extensive use of discriminant functions and factor analysis to analyze a vast amount of
55 craniometric data stands as the exception that proves the rule (Oxnard 1974).

This is not to say that early workers were without computational power, but they had names.

Here is a quote from a letter of Karl Pearson to his funding organization, the Worshipful
Company of Drapers, reprinted in Egon Pearson's 1936 biographical obituary (Pearson 1936). K.
Pearson writes:

60 “This has been altered by the appointment of Dr. Lee as a computator, and by the occasional
payment of additional calculators, e.g., Miss F. Cave of Girton College.”

Today, even ignoring the potential of parallel processing, there is more than enough computing
power available on the desktop along with an accessible panoply of data collection modalities for

Slice: Computational Tools in Forensic Science

physical and forensic anthropologists around the world to realize much of the potential of
65 multivariate tools. However, there are a number of considerations that should be kept in mind in
their development and use.

The first issue to be addressed in the development and use of computational tools is that of the
question being addressed. In this area, applied forensic science has an advantage over many other
fields. The important questions in forensics are often external and clearly stated – What sex was
70 the person represented by the skeletal remains? What was their ancestry? How old were they
when they died? How long have they been dead? And the more general and possibly more
difficult, who were they? With such clear and important questions in hand, one can begin to
develop or select tools appropriate for the question being asked and whatever material might be
available for analysis.

75 Given such a decidedly clear question, the Tool Makers (to borrow a phrase commonly used in
general morphometric parlance), can begin to address ways in which it can be answered for
specific or commonly available data. Here, the importance of the data used to develop a method
cannot be overstated, and its quality and relevance should be given extensive consideration. This
is especially true for quality, as relevance will reveal itself in the course of method development
80 and testing.

There are a number of issues to consider with respect to data quality, but note that sample size is
an issue of quantity, not quality. With sufficient sample size, one is almost guaranteed statistical
significance, but that may or may not be useful in the applied setting. Effect size (Kelley 2012) is
a more appropriate indicator of model utility, e.g., R^2 in a regression problem (Yin and Fan

85 2001), but see the discussion of dimensionality below.

Data quality is an issue of the repeatability of measurements initially within and between observers involved in the method development, but ultimately across end-users. Quantities that cannot be reliably reproduced by different individuals are not particularly helpful for applied use. After repeatability, data accuracy is a primary concern. One must be certain that the data used for
90 method development is correct. It should be checked and rechecked for outliers and simple transcription errors, and here plotting can play a crucial role. To quote the admonition of the late Dr. Leslie L. Marcus, “Plot the hell out of your data!” Plot the raw data. Plot the principle components. Plot the histograms. Plot. Plot. Plot.

I have, on more than one occasion, been asked by both students, established researchers, and
95 editors to look at striking or unusual results from a morphometric analysis. As I tell my students, I refuse. Instead, I ask for the original data. After checking the data to be sure it has been accurately encoded and has no result-determining outliers, I will attempt to reproduce the results. Only then might I have something to say. Often, though, I have found an issue with the data that can explain the anomalous results. Simply being off by one can randomize relationships within
100 unordered data sets, and a group miscoding can shockingly invert otherwise standard results. Check your data.

While sample size, itself, is not particularly relevant for model development where any useful effect should be strong enough to be evident with modest numbers of individuals, it may be problematic relative to data dimensionality. Data dimensionality refers to the number of variables
105 recorded per individual observation. Simple sets of measurements like the length and diameter of

Slice: Computational Tools in Forensic Science

a long bone or the gnathic, cephalic, and other indices present few problems with respect to sample size. With modern data-acquisition techniques and shape-analysis methods data dimensionality can be a substantial consideration. A sample of, say, 1000 individuals should be more than enough to reveal any relationship of morphology to another variable of interest, but

110 consider that if those 1000 individuals have been scanned at even modest resolution, the surfaces may contain an average of 100,000 points each! In three dimensions, this results in 300,000 measured variables. This situation is ripe for overfitting with the ability to completely separate any partitioning of the original data one might specify. Scans of one million three-dimensional points are not uncommon.

115 This leads to the often necessary issue of dimension reduction, the most common method for which is Principle Components Analysis (PCA) (Jackson 2005). It is common for researchers to reduce dimension by taking only projections of the original data onto the first, first three, or first i PCs accounting for x percent of the sample variation. Here it is important to remember that PCA is driven by one thing – directed variance resulting from variable covariance or correlation.

120 Nothing in the machinery of PCA knows about sex or ancestry or age or anything but variance and variable correlation. The first few PCs are sensitive to outliers in the data and even highly variable measurements, either naturally or those just difficult to collect. Plotting the data projections onto the PCs is, in fact, very useful for checking for outliers and highly variable measurements in the data set, but analyzing just the first few or many PCs can completely miss

125 an otherwise strong signal of interest. Again, check your data and its covariance patterns.

Nonetheless, some sort of dimension reduction is often necessary when dealing with complex

Slice: Computational Tools in Forensic Science

data sets, and PCA is a familiar and useful one. One needs to be VERY careful in the selection of components to use and aware of the implications of this choice. For a cautionary note and further discussion of PCA see Bookstein (2017, 2018).

- 130 Another thing to keep in mind is that models are generally optimal for the data used to build the model. As such, comparing predicted values from a model to the data used to build the model can be especially misleading. Instead, one must carry out some sort of validation procedure comparing predictions to known values for individuals completely removed from model construction. One such approach I find particularly useful is Lachenbruch's method of cross-
- 135 validation (Lachenbruch and Mickey 1968) where each individual is removed from the dataset, a model built anew, and that individual's trait predicted from the new model. This gives a realistic indication of the performance of a method, and I even include doing this as an option in, for instance, *3D-ID* (Slice and Ross 2009), so that the user can see for themselves how well the software is expected to perform.
- 140 With respect to Tool Makers, I would suggest a few important attributes of any tool expected to be used by the community. First, the implementation should be platform-independent. This is not that hard to achieve. The R-programming environment (R Core Team 2015) provides many tools for data manipulation and can be downloaded to run on any common platform – Windows, OS X, Linux. Java (<https://www.java.com>) is often the most popular programming language at any
- 145 given time (e.g., <https://www.tiobe.com/tiobe-index/>), has virtual machines available for common operating systems, and development environments allowing for the easy production of simple, intuitive user interfaces. If raw power is required – the previous suggestions are

Slice: Computational Tools in Forensic Science

“interpreted” environments and, hence, slower, than C/C++ are good choices with compilers available for any platform.

150 Secondly, it is critical that software be open-source. That is, the source code should be made freely available for users, but mostly other programmers, to inspect or modify or debug. This is critical, though I have violated this suggestion in the past (but I am trying to do better, e.g., *forAge* (Stoyanova, Algee-Hewitt, and Slice 2015) and *gpsa* (Pomidor and Slice 2015)), because programs distributed as executables, no matter how useful, die with the interest of the
155 programmer or with the programmer, themselves. If source code is available, there is at least a chance that another capable person can take over development and maintenance or find bugs and problems with the original code.

In contrast, the support and maintenance of commercial software is a daunting task, and a lot of money is required to keep on staff a sufficient number of programmers and customer-service
160 representatives. Anything less than a full commercial enterprise behind a package is mostly doomed to failure. However, the money required to support such an enterprise is likely to be more than a relatively small community can provide or even justify. Open-source software of high quality, even from an individual developer, greatly diminishes the chance of later user dependence on an orphaned program.

165 The comments thus far have been mostly directed at Tool Makers. The Tool Users also have a responsibility for the appropriate and efficacious use of the tools provided. Here again, data quality is critical. Even if one has but a single bone, the measurements of which are to be submitted to a classification or prediction program, the collection of those measurements must be

Slice: Computational Tools in Forensic Science

tested and retested to ensure their greatest precision and accuracy. Method developers can help
170 here by providing raw material through which end-users can compare their data collection results
to those assumed by the developer, though this may be practical only for digitally recorded
specimens.

Finally, it is absolutely necessary that the end-user has a sufficient understanding of the methods
being implemented and the data upon which a method or result is based. This is especially
175 critical for medico-legal professionals who might be called upon to give court testimony and be
subject to cross-examination. Any set of appropriately structured numbers submitted to a
program will give some sort of appropriately labeled result. Without understanding the
underlying processing and models, the end-user really has no idea of their validity. This does not
mean that end-users must be able to derive equations or expound on the details of underlying
180 theory, but it does mean they should be able to explain what the program is doing, what
relationships within the data are being exploited, and what the results mean, especially as to the
uncertainty inherent in them. As a general rule, if you cannot explain what you did to a random
person on the street, then you don't understand the method enough to be using it.

The above provides some basic considerations for the development and use of computational
185 models and software in the forensic setting. These should be quite familiar to most involved in
these activities. If they are not, study them carefully and take them to heart.

Literature Cited

Bookstein, F. L. 2017. A method of factor analysis for shape coordinates. *Am J Phys Anthropol*

Slice: Computational Tools in Forensic Science

164(2): 221–45.

190 Bookstein, F. L. 2018. *A Course in Morphometrics for Biologists*. Cambridge University Press.

Cayley, A. 1858. A memoir on the theory of matrices. *Philosophical Transactions of the Royal Society of London* 148: 17–37.

Gower, J. C. 1975. Generalized Procrustes analysis. *Psychometrika* 40(1): 33–51.

Hotelling, H. 1931. The generalization of Student's ratio. *The Annals of Mathematical Statistics*
195 2(3): 360–78.

Howells, W. W. 1973. *Cranial Variation in Man*. Papers of the Peabody Museum of Archaeology and Ethnology, Harvard University, V. 67.

Jackson, J. E. 2005. *A User's Guide to Principal Components*. John Wiley & Sons.

Kelley, K. and K. J. Preacher. 2012. On effect size. *Psychological Methods* 17(2): 137–52.

200 Lachenbruch, P. A. and M. R. Mickey. 1968. Estimation of error rates in discriminant analysis. *Technometrics* 10(1): 1–11.

Mahalanobis, P. C. 1927. Analysis of race mixture in Bengal. *Journal and Proceedings of the Asiatic Society of Bengal* 23:301–333.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proceedings of the National*
205 *Institute of Sciences of India* 2(1): 49–55.

Mosier, C. I. 1939. Determining a simple structure when loadings for certain tests are known.

Slice: Computational Tools in Forensic Science

Psychometrika 4(2): 149–62.

Oxnard, C. E. 1974. Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference among Recent Human Populations. *Am J Phys Anthropol* 41(2): 349–51.

210 Pearson, E. S. 1936. Karl Pearson: an appreciation of some aspects of his life and work.

Biometrika 28(3/4): 193–257.

Pearson, K. 1928. The skull and portraits of Henry Stewart, Lord Darnley, and their bearing on the tragedy of Mary, Queen of Scots. *Biometrika* 20B(1): 1–104.

Pearson, K. and G. M. Morant. 1934. The Wilkinson Head of Oliver Cromwell and its

215 relationship to busts, masks and painted portraits. *Biometrika* 26(3): 1–116.

Pomidor, B. and D. E. Slice. 2016. *gpsa*. <http://morphlab.sc.fsu.edu/software/gpsa/index.html>

R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rohlf F. J. and D. E. Slice. 1990. Extensions of the Procrustes method for the optimal

220 superimposition of landmarks. *Systematic Zoology* 39(1): 40–59.

Slice, D. E. and A. Ross. 2010. *3D-ID: Geometric Morphometric Classification of Crania for Forensic Scientists*. <https://www.ncjrs.gov/pdffiles1/nij/grants/231195.pdf>

Stoyanova, D., B. F. B. Algee-Hewitt, and D. E. Slice. 2015. *forAge*.

<http://morphlab.sc.fsu.edu/software/forAge/index.html>

225 Wilks, S. S. 1946. Sample criteria for testing equality of means, equality of variances, and

Slice: Computational Tools in Forensic Science

equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics* 17(3): 257–81.

Wishart, J. 1928. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 20A(1/2): 32–52.

230 Yin, P. and X. Fan. 2001. Estimating R^2 shrinkage in multiple regression: a comparison of different analytical methods. *The Journal of Experimental Education* 69(2): 203–24.