The Pennsylvania State University The Graduate School College of the Liberal Arts

IMPROVED SKELETAL AGE-AT-DEATH ESTIMATION AND ITS IMPACT ON ARCHAEOLOGICAL ANALYSES

A Dissertation in

Anthropology

by

Sara Marie Getz

© 2017 Sara Marie Getz

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

May 2017

The dissertation of Sara Marie Getz was reviewed and approved* by the following:

George R. Milner
Professor of Anthropology
Dissertation Advisor
Chair of Committee

James W. Wood Professor of Anthropology and Demography

Joan T. Richtsmeier Professor of Anthropology

Neil A. Sharkey Professor of Kinesiology, Orthopedics, and Rehabilitation

Jesper L. Boldsen Professor, Institute of Forensic Medicine University of Southern Denmark Special Member

Kenneth Hirth
Professor of Anthropology
Director of Graduate Studies

^{*} Signatures are on file in the Graduate School.

ABSTRACT

Skeletons are the most direct way to study long-term trends in longevity, mortality patterns, and disease experience for much of human existence. Adequate age estimates can be produced for children and young adults, but such estimates for most of adulthood remain elusive. This fourphase dissertation is a large-scale proof-of-concept that accurate and precise age estimates can be produced for all of adulthood without the two most widely used skeletal indicators—the pubic symphysis and auricular surface. This work builds on over two decades of research by an international research team and an existing age-estimation method called Transition Analysis (TA). In Phase 1, more than 200 trait variants were investigated to identify and refine a set of age-informative features throughout the skeleton, and primary reference data were collected from four collections of modern, known-age North American skeletons (N=1010). In Phase 2, a simplified procedure based on existing TA was developed to produce age estimates from Phase 1 reference data. In Phase 3, standard age-estimation methods and new TA were applied to additional known-age samples—one modern and one historical. In both samples, the estimates produced by new TA have similar accuracy to traditional methods, but the ranges are, on average, half as wide and show essentially no systematic point estimate bias. In Phase 4, traditional methods and new TA were applied to two Danish archaeological samples. Comparing the Phase 3 and 4 samples reveals that each traditional method produces a characteristic pattern of adult mortality that is practically independent of the age distribution of the sample. Thus, all mortality profiles generated for past populations using traditional techniques should be viewed with critical skepticism. In contrast, new TA produces mortality distributions that more closely approximate reality, including details that could not be detected using traditional techniques. This dissertation, in conjunction with a larger NIJ-funded research project using the same approach for geographically diverse modern populations, provides every indication that the new TA procedure may become the new gold standard for adult age estimation.

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	xiii
ACKNOWLEDGMENTS	xiv
CHAPTER 1: INTRODUCTION	
Dissertation Overview)
collections	
Phase 2: Collaborative trait refinement, error testing, and method development	
Phase 3: Validation studyPhase 4: Archaeological application	
Dissertation Structure	
CHAPTER 2: AGING OF THE ADULT SKELETON—THEORY & PRACTICE	
Aging and the Adult Skeleton	
Chronological versus biological age	
Activity, aging, and disease	
Decline in the rate of increase of age-specific mortality rates	
Traditional Approaches to Adult Skeletal Age-Estimation	
Skeletal samples	
Applying age-estimation methods to skeletal samples	
Significance of accurate and precise age estimates	
Moving Forward: This Dissertation & the NIJ Project	17 18
CHAPTER 3: PHASE 1—PRELIMINARY TRAIT IDENTIFICATION & REFINEMENT USIN NORTH AMERICAN SAMPLES	19
Identifying and Defining New Age-Informative Skeletal Features	20
Using logistic regression to select new age-informative skeletal traits	
Data Collection	
Pilot study & preliminary data collection	
Primary data collection	
Skeletal collections	
Trait Selection Trait examples and interpretation	
Final Phase 1 Traits	
CHAPTER 4: PHASE 2—COLLABORATIVE TRAIT INVESTIGATION & REVISED	
TRANSITION ANALYSIS METHOD DEVELOPMENT	
NIJ Project Objectives	
The NIJ Project and Dissertation Phase 2	
Establishing a baseline for method improvement	
NIJ Data Collection	
Skeletal samples	
Identifying and refining skeletal traits	
Establishing trait age distributions	4/

Age-At-Death Estimates Using Multiple Features	
Maximum likelihood estimates	
Preliminary age estimates for Bass Collection skeletons	
Age intervals for maximum likelihood estimates	
Correlated traits	
The New Transition Analysis Method Software	59
Software development	59
Sex and population-specific reference samples	
Prior distributions	
Trait scoring manual	63
CHAPTER 5: PHASE 3—TESTING REVISED TRANSITION ANALYSIS OF	
AGE SAMPLES	
Phase 3: Method Validation	65
Phase 3 samples	
Establishing a baseline for method improvement	
The New TA Procedure	
Establishing probability density functions for skeletal features	
Combining features to estimate age	
Correlations between traits	
Evaluating Variations of the New TA Procedure: The Athens Sample	
Using a reduced feature set to estimate age	
Choice of reference sample: a cautionary note	
Performance of new TA relative to commonly used methods	
Evaluation of New TA on a Historical Sample: St. Bride's Crypt	
The effect of skeletal preservation	
Accuracy and precision	
Point estimates and age-estimation bias	
Point estimates and age-at-death distributions	
Phase 4: An Archaeological Application of the New TA Procedure	100
CHAPTER 6: PHASE 4—ARCHAEOLOGICAL APPLICATION: THE DANI	
HORSENS IN THE LATE MEDIEVAL & EARLY MODERN PERIOD	
Phase 4 Samples	
HOM 1649—Ole Worms Gade (Ole Worms Street)	
HOM 1272—Horsens Klosterkirke (Monastery Church)	104
Methodological Considerations	
Traits to include	
The effect of too few traits	
Preservation of commonly used adult age indicators	
The performance of standard age indicators on past populations	
Standard Methods & Implications for Paleodemographic Applications	
Three standard age-estimation methods	
Existing TA, New TA, and experienced-based estimates	
Assessing Mortality in the Danish City of Horsens	
Conclusions	117

CHAPTER 7: DISCUSSION, CONCLUSIONS, & FUTURE DIRECTIONS	118
Discussion	120
Trait selection and data collection procedures	
Correlated traits	
Sex- and population-specific reference standards	
Prior distributions	
Conclusions	
Ongoing Work	123
Future Directions	124
APPENDIX A: PUBLICATIONS RELATED TO ADULT SKELETAL AGE ESTIMAT MACROSCOPIC FEATURES OF THE CRANIAL SUTURES, RIBS, PUBIC SYMPI SACRO-ILIAC JOINT	HYSIS, &
APPENDIX B: PHASE 1—TRANSITION CURVES FOR SELECTED FEATURES	130
APPENDIX C: TRANSITION ANALYSIS EQUATIONS	152
APPENDIX D: TRAIT DEFINITIONS EXTRACTED FROM THE CURRENT NIJ SCO	
APPENDIX E: CORRELATION MATRICES FOR PHASE 3 TRAITS	190
REFERENCES	199

LIST OF FIGURES

Figure 2.0. Location of the pubic symphysis and phases of the Suchey-Brooks adult age-estimation method (Brooks & Suchey, 1990). Early and late examples of each phase are shown, along with the associated mean (gray) and range (range) for each. Example figures drawn by George Milner from Tocheri et al. (n.d.) and Diane France casts
Figure 2.1. A schematic overview of the factors in archaeological investigations that may influence the skeletal sample that is ultimately available for analysis
Figure 3.0. Example logistic curves with transitions in differ parts of the adult lifespan Fitted logistic models (solid black lines) with approximate 95% confidence intervals (dashed black lines) and estimated ages-at-transition (vertical dotted lines) based on simulated trait data (black dots)24
Figure 3.1. Example transition plots from Figure 3.0 (a1, b1, and c1), shown with the logistic curves for both the presence and absence of a trait (a2, b2, and c2). The estimated age-at-transition (top row, vertical dotted line) is the point at which there is an equal probability of a trait being absent (bottom row, dashed curve) and present (bottom row, solid curve)
Figure 3.2. Schematic of the trait identification and refinement process in Phase 1, beginning with trait identification and definition (upper left) and ending when a suite of traits that provides information across the entire adult lifespan has been identified (upper right). 28
Figure 3.3. Age distributions of males and females in the Phase 1 primary reference sample32
Figure 3.4. Age distribution, by sex, of the individuals evaluated from each skeletal collection and death cohort
Figure 3.5. Key features of the transition plots used to evaluate Phase 1 traits34
Figure 3.6. Transition plots for irregular bony growth on the margin of the lesser tubercle of the humerus.
Figure 3.7. Transition plots for the presence of exostoses on the anterior inferior iliac spine (AIIS)37
Figure 3.8. Transition plots for the presence of candlewax on at least one cervical vertebra38
Figure 3.9. (a) Distribution of the approximately 100 features investigated in Phase 1 and (b) the location of the final 45 features identified as potentially informative for adult age estimates39
Figure 4.0. Age estimates using standard procedures for the sacroiliac joint and pubic symphysis and existing TA based on both the pelvis and cranium. Dots indicate published central tendencies and plus signs indicate the lower bounds of terminal open-ended intervals (males in blue and females in red). Figure created by Getz with Milner data (Bass Collection, N=234); data previously presented in (Milner, Boldsen, Ousley, Weise, et al., 2016)
Figure 4.1. Existing methods produce estimates along the line from low accuracy with high precision (lower left) to high accuracy with low precision (upper right). Figure created by Getz with Milner data (Bass Collection, N=234); presented in (Milner, Boldsen, Ousley, Weise, et al., 2016)45
Figure 4.2. Comparison of the GAMs generated with the smallest (10 ⁻⁶ , black line) and largest (10 ³ , grey line) smoothing parameters for a) lateral clavicle macroporosity, b) sternal dorsal ridges, and c) thoracic vertebrae candlewax. Data from Getz Phase 1 primary reference sample
Figure 4.3. Comparison of the logistic (grey line) and generalized additive model (black line) for the a) profile of ribs three through ten, b) humerus lateral epicondyle, and c) lumbar vertebral lipping, with the greatest divergence occurring at the ends of the age distribution. Data from Getz Phase 1 primary reference sample
Figure 4.4. Two features for which the choice of smoothing parameter has essentially no effect. Logistical curve (solid grey line) and ten different generalized additive models with different smoothing parameters (colored dotted lines) for medial clavicle (left) and bone growth on the lateral margin of the lesser tubercle of the humerus (right). Data from Getz Phase 1 primary reference sample

middle age likely reflect slight fluctuations in the sample age distribution, but are unlikely to be related to underlying biological processes. Logistic curve (grey solid line) and ten GAMs (colored dotted lines) are shown for textural changes on the medial epicondyle of the humerus (left) and lipping of the lumbar vertebrae (right). Data from Getz Phase 1 primary reference sample
Figure 4.6. Examples of traits where the smoothing parameter has a significant effect on the shape of the fitted curves, especially at the oldest ages. Models shown for hyperostosis frontalis interna (HFI) (left) and extremely thin ribs (right). Data from Getz Phase 1 primary reference sample
Figure 4.7. Logistic (grey line) and generalized additive models with an automatically chosen smoothing parameter (black line) for candlewax (DISH) in the thoracic vertebrae (left) and irregular ossifications on the profile of ribs three through ten (right). Data from Getz Phase 1 primary reference sample
Figure 4.8. Logistic (grey line) and generalized additive models with an automatically chosen smoothing parameter (black line) for superior-anterior sacro-iliac joint fusion (left) and development of the public symphyseal collar (right). The bumps around 60 years of age may possibly indicate selective mortality and are smoothed over by the logistic model. Data from Getz Phase 1 primary reference sample
Figure 4.9. (a) Maximum likelihood age estimates produced using existing TA based only on cranial and pelvis traits, and (b) the new procedure using features throughout the skeleton for the same sample (males in blue and females in red). Identify line shown for comparison. Age estimates generated by Boldsen using Milner data (Bass Collection, N=234) and figure created by Getz; data presented in (Milner, Boldsen, Ousley, Getz, et al., 2016)
Figure 4.10. Comparison of new TA maximum likelihood estimates (males in blue and females in red) for the same sample based on (a) all features showing age-related patterns and (b) the same feature set with traits showing variation in the sex-specific GAM curves eliminated. Identify line shown for comparison. Age estimates generated by Boldsen using Milner data (Bass Collection, N=234) and figure created by Getz
Figure 4.11. Comparison of maximum likelihood age estimates produced for the same individuals using a logistic model (GML), a generalized additive model with an automatically selected smoothing parameter (GAM1), and an alternative GAM with a hand-selected smoothing parameter (GAM2), all based on a combined sex reference sample. Identity line shown for reference. Getz data, Athens Collection, N=201.
Figure 4.12. Comparison of the maximum likelihood age estimates produced for the same individuals using sex-specific models. Identity line shown for reference. Getz data, Athens Collection, N=20156
Figure 4.13. A log-likelihood function based on 33 skeletal features showing a) a horizontal dashed line 1.9208 units below the peak of the distribution and b) the 95% interval based on the intersection of the horizontal line with the probability curve (vertical dotted lines). Open circle represents the true age of a female from the Athens Collection
Figure 4.14. Probability curves and bar plot output from the existing TA program (ver. 2.1.041. The curves and bars show the age range produced by each of the areas of the skeleton individually (auricular surface—green, pubic symphysis—magenta, cranial sutures—red) and the combined age estimate using a uniform (black) or archaeological (blue) prior distribution
Figure 4.15. Maximum likelihood age estimates for the same individuals using three models generated from combined (left) and sex-specific (bottom) reference samples. Identity line shown for reference. Getz data, Athens Collection, N=201
Figure 4.16. Comparison between maximum likelihood age estimates produced using the uniform and archaeological prior distributions in existing TA. The identify line is shown for reference. Getz Data: Athens Collection, N=201
Figure 5.0. Age distributions of individuals evaluated from the Athens and St. Bride's samples (top row) and the same individuals separated by sex (bottom row, males in blue and females in red)

Figure 5.1. Age estimates generated by Milner (c) and Getz (d) for different samples (a,b) using commonly used pubic symphysis (black circles) and sacroiliac joint (grey circles) methods, existing TA and experience-based estimates (white circles) showing the same pattern for different samples
Figure 5.2. Eighteen variations of how probabilities can be calculated for traits from the reference sample collected in Phase 1. Only fifteen are evaluated here because there are too few females in death cohort 1 (grey boxes) to allow sex-specific probability curves to be generated
Figure 5.3. Each of the four point clusters shows the average accuracy and precision of age estimates for the entire Athens sample using trait probabilities from 12 slightly different model variations (shown in Figure 5.2). Differences between clusters are the result of the number of features (circles/triangles) and the prediction interval (filled/open) used to calculate the maximum likelihood estimate and its associated range.
Figure 5.4. Results of the same method variations shown in Figure 5.3, identified by model choice reference sample, and whether or not sex-specific probabilities were used. Variations using a combined non-sex-specific reference sample resulted in the best compromise between accuracy and precision. No other consistent patterns are present
Figure 5.5 Four traits identified as the combination of features that provided the best coverage of the entire adult age range: a) medial clavicle epiphysis fusion, b) scapula glenoid fossa margin, c) humerus lesser tubercle bumps, and d) humerus weight
Figure 5.6. Age distributions of the known (a1,a2) and estimated ages of the "uniform" test sample (N=300) produced using the iteratively-selected features (b1,b2), shown in approximately one-year (a1,b1) and five-year (a2,b2) categories
Figure 5.7. Comparison of new TA method variations using the logistic probabilities from different numbers of features. Circles (all 40 traits) and squares (30 traits) are the same data shown for both 95% and 99% prediction intervals. These are compared to the results using smaller samples of hand-selected (14, diamonds) and iteratively selected features (21, triangles). Athens Collection, N=201
Figure 5.8. Comparison of ages generated with logistic model trait probabilities based on different sets of features and reference samples. Using a reference sample that is a poor fit for the test samples results in wider age ranges, lower precision, with reduced accuracy. Athens Collection, N=20182
Figure 5.9. Direct comparison of the accuracy and precision of age estimates generated using variations of the new TA procedure (black symbols), commonly used methods for the pubic symphysis and auricular surface, and experience based estimates. Athens Collection, N=20183
Figure 5.10. RTO coefficients and associated confidence intervals for age estimates produced for the Athens sample (N=201) using standard methods (1-7), observer experience (8), and variations of the new TA procedure (9-40). Confidence intervals that include one (dotted box) indicate that the estimates produced are collectively unbiased.
Figure 5.11. Age estimates produced by three commonly used techniques: (a) (Lovejoy, Meindle Pryzbeck, et al., 1985), (b) (Brooks & Suchey, 1990; Suchey et al., 1988), and (c) (Buckberry & Chamberlain, 2002) with a fitted linear regression (dotted line), regression coefficient, and associated confidence interval. Identify line (solid line) shown for comparison. Athens Collections, N=20185
Figure 5.12. Age estimates produced by an experienced observer (Getz), existing TA (Boldsen et al., 2002), and four variations of the new TA procedure using logistic regression probabilities from different numbers of features. The identity line (solid line) and fitted regression through the intercept (dotted line) are shown for comparison. Athens Collection, N=201
Figure 5.13. Age ranges produced by an experienced observer, existing TA, several standard methods, and the two variations of the new TA procedure. Black dots indicate the known age of each individual, grey lines indicate the age ranges estimated by each method. Athens Collection, N=201
Figure 5.14. Percentage of the St. Brides crypt sample (N=168) that could be aged using methods for the pubic symphysis (light grey), auricular surface (dark grey), and those that use features from multiple skeletal elements (experience-based estimates, existing TA, and new TA) (black)

Figure 5.15. Distribution of the number of scorable features of each individual in the St. Brides sample (N=168)
Figure 5.16. Comparison of age estimates produced for the Athens sample (N=201) using combined reference sample logistic probabilities from different numbers of features92
Figure 5.17. Relationship between the number of features used to estimate age and the length of the interval produced using (a) combined or (b) time period-specific trait probabilities and a 95% prediction interval. St. Bride's Crypt sample (N=168)
Figure 5.18. Accuracy and precision of age estimates generated using new TA procedure variations (dotted box), commonly used methods for the pubic symphysis and auricular surface (open symbols), and experience-based estimates (asterisk) for the St. Bride's Crypt sample (N=168)93
Figure 5.19. Comparison of the age ranges produced by an experienced observer, existing TA, severa standard methods, and two variations of the new TA procedure for the St. Bride's Crypt sample (N=168). Black dots indicate the known age of each individual and grey lines indicate the range estimated by each method
Figure 5.20 Ages estimates produced for the St. Bride's sample (N=168) by two commonly used methods (Buckberry & Chamberlain, 2002) an experienced observer, existing TA (middle row), and two variations of the new TA procedure using logistic regression probabilities for different numbers of features (bottom row). The identity line (solid line) and regression fitted through the origin (dotted line) are shown for comparison
Figure 5.21 Known (solid line) and estimated (dashed line) age-at-death distributions for the St. Bride's crypt sample produced using standard techniques (left column), existing TA, observer experience, and new TA98
Figure 6.0. Location of Phase 4 archaeological samples from Horsens, Denmark at different scales: (a) Northern Europe and Scandinavia, (b) Denmark, (c) city of Horsens, and (d) approximate excavation locations. The rectangular box on each map (a-c) indicate the area enlarged in the next frame. In frame (d), the highlighted areas are the two sites—(1) Klosterkirke and (2) Ole Worms Gade. The trapezoida box near (2) indicates the likely extent of non-excavated portion of the cemetery. Maps modified from (Google, 2016)
Figure 6.1. Distribution of the number of features available for estimating age for the individuals in each sample—(a) HOM 1649: Ole Worms Gade (N=317) and (b) HOM 1272: Klosterkirke (N=168)—when either 30 or 40 traits were possible. Note the difference in the y-scale between the two charts 105
Figure 6.2. Getz experience-based estimates compared to maximum likelihood estimates generated from logistic trait probabilities for a maximum of 40 features for the St. Bride's sample (N=168) and Klosterkirke (HOM 1272, N=166). The same data are shown in the panels for each sample; estimates produced using fewer than 5 or 15 features are shown in red
Figure 6.3. Percentage of the HOM 1649 (Ole Worms Gade, N=317) and HOM 1272 (Klosterkirke, N=166) samples that could be aged using methods for the pubic symphysis (Suchey-Brooks), auricular surface (Lovejoy and B&C), and those that use features from multiple skeletal elements (experience existing TA, and new TA)
Figure 6.4. Comparison of the age ranges produced by the two best-performing standard methods for the pubic symphysis and auricular surface, existing TA, and the new TA procedure using 40 features for the documented St. Bride's sample and the Klosterkirke archaeological sample. Black dots in each graph show Getz experience-based estimates, while the grey lines indicate the range estimated for these individuals by each method. 110
Figure 6.5. Top row: The documented (solid line) and estimated (dashed line) age-at-death distributions for the St. Bride's sample do not resemble one another. Bottom row: Estimated age distributions for the St. Bride's and HOM 1272 (Klosterkirke) samples show similar method-specific patterns112

Figure 6.6. Comparison of the mortality distributions produced for the Phase 3 know-age samples (Athens and St. Bride's) and Phase 4 archaeological samples (HOM1649 and HOM 1672) based on age estimates from three standard methods113
Figure 6.7. Comparison of the mortality distributions produced for the two Phase 3 known-age collections (Athens and St. Bride's) and the two Phase 4 archaeological samples (HOM1649 and HOM 1672) using existing TA, Getz experience-based estimates, and new TA114
Figure 6.8. Age-at-death distributions for the individuals of the Ole Worms Gade (HOM 1649, solid line and Klosterkirke (HOM 1272, dashed line) samples based on estimates from commonly used techniques (left column), existing TA, observer experience, and new TA (right column)
Figure 7.0: Timeline for this dissertation and the collaborative work conducted as part of the NIJ project
Figure B1.0. Key components of the transition plots used to evaluate Phase 1 traits130
Figure B1.1. Transition plots for HFI, parietal depressions, and occipital condyle lipping13
Figure B1.2. Transition plots for first rib fusion, R2 sternal end edge profile, and R3–10 sternal end edge profile
Figure B1.3. Transition plots for R3–10 body thickness ("shingle ribs") and sternal central dorsal ridge
Figure B1.4. Transition plots for cervical, thoracic, and lumbar vertebral lipping
Figure B1.5. Transition plots for cervical, thoracic, and lumbar vertebral candlewax138
Figure B1.6. Transition plots for C1 articular facet margin lipping and C1 articulation facet surface eburnation136
Figure B1.7. Transition plots for L5 superior margin shape137
Figure B1.8. Transition plots for L5 inferior margin shape138
Figure B1.9. Transition plots for S1 superior margin shape139
Figure B1.10. Transition plots for S1-2 fusion and superior anterior sacroiliac joint fusion140
Figure B1.11. Transition plots for clavicle medial epiphysis fusion14
Figure B1.12. Transition plots for clavicle medial epiphysis irregular bone growth and clavicle latera articulation surface macroporosity142
Figure B1.13. Transition plots for scapula glenoid fossa lipping143
Figure B1.14. Transition plots for humerus medial epicondyle, humerus lateral epicondyle, and humerus lesser tubercle margin lipping144
Figure B1.15. Transition plots for humerus lesser tubercle anterior surface bone growth, radius tuberosity medial crest, and trapezium lipping145
Figure B1.16. Transition plots for femoral fovea margin lipping and femoral head surface bone growth
Figure B1.17. Transition plots for femur trochanteric fossa exostoses and femur trochanteric fossa media surface exostoses
Figure B1.18. Transition plots for acetabulum posterior margin lipping, acetabulum posterior articulation surface bone growth, and pubic symphysis ventral collar148
Figure B1.19. Transition plots for ischium superior margin spur, ischium surface irregular bone growth and AIIS exostoses
Figure B1.20. Transition plots for the subjective weight of the humerus, tibia, and innominate150
Figure B1.21. Transition plots for the subjective weight of the calcaneus15

Figure D1.0. NIJ trait manual cover page	154
Figure D1.1. Parietal depression description and examples (page 1 of 2)	155
Figure D1.2. Parietal depression description and examples (page 2 of 2)	156
Figure D1.3. C1 articular facet surface margin lipping description and examples	157
Figure D1.4. C1 articular facet surface eburnation description and examples	158
Figure D1.5. L5 margin shape description and examples.	159
Figure D1.6. S1 margin shape description and examples	160
Figure D1.7. S1-2 fusion description and examples.	161
Figure D1.8. Cervical, thoracic, and lumbar vertebral lipping description and examples	162
Figure D1.9. Cervical, thoracic, and lumbar vertebral candlewax description and examples	163
Figure D1.10. Sternum central dorsal ridge description and examples (page 1 of 2)	164
Figure D1.11. Sternum central dorsal ridge (page 2 of 2) description and examples (page 2 of 2)	165
Figure D1.12. R2 and R3-10 sternal end rim profiles description and examples	166
Figure D1.13. R3-10 rib body thickness (shingle ribs) description and examples	167
Figure D1.14. Clavicle medial epiphysis fusion description and examples	168
Figure D1.15. Clavicle medial epiphysis gravel description and examples	169
Figure D1.16. Clavicle medial macroporosity description and examples.	170
Figure D1.17. Clavicle lateral macroporosity description and examples.	171
Figure D1.18. Scapula glenoid fossa raised border and lipping description and examples	172
Figure D1.19. Humerus lesser tubercle anterior surface bumps description and examples	173
Figure D1.20. Humerus lesser tubercle margin shape description and examples	174
Figure D1.21. Humerus medial epicondyle description and examples	175
Figure D1.22. Humerus lateral epicondyle description and examples.	176
Figure D1.23. Radial tuberosity medial crest description and examples	177
Figure D1.24. Trapezium lipping description and examples	178
Figure D1.25. Femur fovea margin lipping description and examples.	179
Figure D1.26. Femur head surface extra bone growth description and examples	180
Figure D1.27. Femur trochanteric fossa exostoses description and examples.	181
Figure D1.28. Femur medial trochanteric fossa exostoses description and examples (page 1 of 2)	182
Figure D1.29. Femur medial trochanteric fossa exostoses description and examples (page 2 of 2)	183
Figure D1.30. Superior-anterior sacroiliac joint fusion description and examples	184
Figure D1.31. Ilium AIIS exostoses description and examples.	185
Figure D1.32. Acetabulum posterior margin lipping description and examples	186
Figure D1.33. Acetabulum articular surface extra bone growth description and examples	187
Figure D1.34. Ischial tuberosity bumps description and examples.	188
Figure D1.35. Subjective bone weight descriptions	189
Figure E1.0 Locations of the features in Phase 3 trait correlation matrices (Tables E1.0-E1.7)	190

LIST OF TABLES

Table 3.0. North American skeletal collections used in Phase 1	26
Table 3.1. Summary of the Phase 1 trait identification & refinement process	28
Table 3.2. Phase 1 primary reference sample death cohort information	32
Table 3.3. Phase 1 primary reference sample trait analysis sample sizes	35
Table 4.0. Skeletal collections evaluated by the NIJ project in Phase 2	46
Table 4.1. Summary of Phase 2 trait refinement and data collection process	47
Table 5.0. Skeletal collections used to test the new TA procedure in Phase 3	66
Table 5.1 Correlation matrix for the subjective weights of four bones	73
Table 5.2 Correlation matrix for four traits on the humerus	73
Table 5.3 Correlation matrix for traits based on thin and light bones	74
Table 5.4 Correlation matrix for lipping and candlewax in three areas of the vertebral column	74
Table 5.5 Correlation matrix for several traits based on excessive bone formation or resorption	74
Table 5.6 Combinations of traits tested in variations of the new TA procedure	80
Table 6.0. Archaeological collections used in Phase 4	103
Table 7.0. Samples collected in each research phase	119
Table C1.0. Steps to calculate probability from the logistic model	152
Table E1.0. Correlation matrix for traits 1-6	191
Table E1.1. Correlation matrix for traits 7-11	192
Table E1.2. Correlation matrix for traits 12-16	193
Table E1.3. Correlation matrix for traits 17-22	194
Table E1.4. Correlation matrix for traits 23-27	195
Table E1.5. Correlation matrix for traits 28-32	196
Table E1.6. Correlation matrix for traits 33-36	197
Table E1.7 Correlation matrix for traits 37.40	198

ACKNOWLEDGMENTS

First and foremost, this project would not have been possible without the two decades of prior work by my advisor, George Milner and, our colleague, Jesper Boldsen. To them, thank you for daring to try something different and setting out this adventure. In particular, I want to thank George, who took the time and risk in 2006 to mentor me—a then inexperienced undergraduate student working on her honors thesis—and set me on the path that would ultimately lead to this dissertation. Additionally, I must also thank the members of the research team and other individuals who have contributed to the foundations of this work—Steve Ousley Peter Tarp, Svenja Weise, Jutta Gampe, James Wood, and Lyle Konigsberg.

This work would not have been possible without the support of the National Science Foundation (NSF) that funded this dissertation, and the National Institute of Justice (NIJ) that funded the larger research team project.

- NSF Award #1455810: Doctoral Dissertation Research: Improved Skeletal Age-at-Death Estimation and Its Impact on Archaeological Analyses (GR Milner, advisor)
- NIJ Award # 2014-DN-BX-K007: Adult Age Estimated From New Skeletal Traits And Enhanced Computer-Based Transition Analysis (PI: George Milner, Co-PIs: Jesper Boldsen and Steve Ousley)

Thank you to the following individuals and institutions that provided access to the skeletal collections that made this work possible: ¹locations of NSF and ²NIJ-funded research that directly contributed to this project and ³locations of essential pilot studies and the contributing work conducted over many years by members of the research team.

- Joshua Barker—Dept. of Anthropology, University of Toronto¹
- Jelena Bekvalac—Centre for Human Bioarchaeology, Museum of London¹
- Jesper Boldsen—The Department of Anthropology (ADBOU), Institute of Forensic Medicine, University of Southern Denmark^{1,2,3}
- Dennis Dirkmaat and Steven Symes—Dept. of Applied Forensic Sciences, Mercyhurst University^{2,3}
- Heather Edgar—Maxwell Museum, University of New Mexico¹
- Susana Garcia—Lisbon Natural History Museum, University of Lisbon²
- David Hunt—National Museum of Natural History, Smithsonian Institution³
- Lee Jantz and Dawnie Steadman—Dept. of Anthropology, University of Tennessee^{1,2,3}
- Erika L'Abbé and Marius Loots—Dept. of Anatomy, The University of Pretoria²
- Pasuk Mahakkanukrauh—Forensic Osteology Research Center, University of Chiang Mai²
- Shirley Schermer—lowa Office of the State Archaeologist¹
- Ana Luísa Santos—University of Coimbra³
- St. Bride's Church (London, UK)¹
- Efstratios Valakos and Panagiota Papazafiri—Dept. of Biology, University of Athens¹
- Michael D. Wiant and Terrance Martin—Illinois State Museum, Norris Farms Skeletal Collection³

This work was also helped by the staff and students who facilitated day-to-day operations at these collections—Rebecca Wilson and Heli Maijanen (Bass Collection), Kathy David (J.C.B. Grant Collection), Lexi O'Donnell and Anna Medendorp (Maxwell Museum Collection), Robin Lillie (Office of the State Archaeologist), and Christina Nicholas (graduate student, University of Iowa). Without your assistance and good humor, this work would not have been possible. A particular thanks goes to the graduate students who opened their doors to help stretch my grant money as far as possible.

This list would not be complete without thanking my family. To my mom—you taught me patience, creativity, and that being polite can go a long way toward meeting your goals. To my dad—you taught me confidence, persistence, and that almost any project can benefit from a list and a binder. And to Abby—you taught me that you can be successful without always being serious and that making time for life should be an important priority. Finally, to my fiancé Patrick—thank you for your patience and support during the nearly a year that I was away collecting these data and the months of craziness it took to put this all together. We made it and our relationship, cats, and houseplants, are now stronger than they were before. We have come a long way from that first conversation when we realized that our research is more similar than anyone would imagine. Thank you for encouraging me to see the possibilities in my data and patiently helping with the code to bring them to life. I look forward to our life of learning together.

Until easier, yet sophisticated, laboratory methods become available the ageing of skeletal remains will always remain a matter of inspired estimation and a little good fortune. Alan A. Watson 1

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Winston Churchill ²

¹ *J For Sci Soc* 14 (3): p.212 ² "The End of the Beginning" - November 10, 1942

CHAPTER 1: INTRODUCTION

Estimating how old an individual was when death occurred is one of the essential steps of most skeletal analyses. This is true whether the ultimate goal is to assess a single individual, such as an isolated grave or in a forensic case, or a larger skeletal sample, as is needed in paleodemography and paleoepidemiology. In forensic contexts, skeletal evidence is often critical to investigations of homicides, missing persons, migrant border deaths, and war crimes. Archaeologically, skeletal remains are one of the only, and most direct, ways of approaching comparative studies of longevity, mortality patterns, social organization, life history courses, disease experience, and population growth for most of human existence. Although forensic and archaeological investigations differ in scope and potential legal ramifications, both hinge on skeletal age-at-death estimates that are sufficiently accurate and precise to allow interpretations to be presented with a reasonable and quantifiable degree of confidence.

A century of method development and refinement has resulted in techniques that can produce reasonable estimates of age for infants, children, and adolescents (Lewis & Flavel, 2006). Unfortunately, age estimation for much of the adult lifespan, particularly for individuals over the age of 50, remains problematic (Falys & Lewis, 2011; Milner & Boldsen, 2012b). Over the past three decades, the evidence indicating that existing methods are not only inadequate, but likely detrimental to many investigations has greatly expanded (Buckberry, 2015). However, age estimates are often presented without full disclosure, or sometimes even acknowledgement, of their flaws or the limitations of what can actually be said from skeletons. Most critically, existing methods consistently produce biased age estimates and little can be said with reasonable confidence for individuals in the latter half of the adult lifespan. Typically, methods overestimate age for young adults, increasingly underestimate it after around age 40, and lump a large fraction of the population into a terminal open-ended category (e.g., 50+). Attempts have been made to mitigate the well-documented problems with standard methods by using new ways of subdividing skeletal features and more complex statistical procedures (M. F. Anderson, Anderson, & Wescott, 2010; Buckberry & Chamberlain, 2002; Buk, Kordik, Bruzek, Schmitt, & Snorek, 2012). These applications, however, have only demonstrated that the most commonly used skeletal features exhibit insufficient age-related variation after middle age to generate improved age estimates (Falys & Lewis, 2011; Milner & Boldsen, 2012c).

In forensic settings, the stakes for obtaining accurate and precise age estimates are high because age-estimation error can have serious ramifications for medicolegal investigations. Accordingly, existing age-estimation methods have been subject to increased scrutiny, and medicolegal standards for method development and evaluation are becoming increasingly stringent (Garvin & Passalacqua, 2012; National Research Council, 2009). Although the archaeological consequences of age-estimation error may seem less drastic by comparison, a debate has simmered for over 30 years regarding the appropriateness of existing methods for archaeological samples and the impact of age-estimation error on our interpretations of the past (Bocquet-Appel & Masset, 1982, 1985; Buikstra & Konigsberg, 1985; Konigsberg & Frankenberg, 1994; Milner & Boldsen, 2012a; Van Gerven & Armelagos, 1983).

One issue highlighted in this debate is whether individuals in prehistory lived beyond 50 years of age and, if they did, how many (Chamberlain, 2006). For individuals truly under 50 years of age existing methods work reasonably well, at least in comparison to their performance for older adults. So, if most or all individuals in a prehistoric population lived to less than age 50, it is possible that the data used to generate interpretations of the past are relatively accurate. However, if some individuals lived to more advanced ages, they could not be confidently

identified using any existing age-estimation technique. Thus, our current understanding of changes in age-specific mortality rates, social structure, and work capacity may be seriously flawed by age-estimation errors. Regardless of what side of the debate one takes, it is impossible to estimate the fraction of the population that lived to over 50 years of age with a reasonable degree of confidence from skeletal samples using traditional techniques. This dissertation contributes to resolving this long-standing archaeological debate, as well as to all endeavors that rely on accurate skeletal data.

Dissertation Overview

I present a new method for adult age estimation based on a previously developed statistical framework, a validation study, and the first archaeological application of the new procedure. The foundation for this work is almost two decades of ongoing research by an international team focused on the complex theoretical, statistical, and methodological components of the adult age-estimation problem (Boldsen et al., 2002; Milner and Boldsen, 2012). Beginning in 1996, Dr. George Milner and Dr. Jesper Boldsen began collaborative work on the skeletal and statistical components of adult age estimation and, in 2002, published a new age-estimation method called Transition Analysis (TA), along with collaborators Dr. James Wood and Dr. Lyle Konigsberg. Although this method produces significantly improved adult age estimates, the performance of the technique is hindered by its reliance on the cranium and pelvic joints (Milner & Boldsen, 2012c). These features, which are the basis of many commonly used age-estimation techniques, experience insufficient change after around age 50 to produce adequate age estimates in the latter half of the lifespan. Therefore, for the next decade, an informal research team continued work on this project as time and money permitted.

As this work expanded in both scope and complexity, the research team also grew in order to tackle the osteological, mathematical, and technological challenges associated with developing, testing, and disseminating a new method of adult age estimation. The team currently includes individuals from four institutions: The Pennsylvania State University (George R. Milner and Sara M. Getz); The University of Southern Denmark (Jesper L. Boldsen, Svenja Weise, and Peter Tarp); The Max Planck Institute for Demographic Research—Rostock, Germany (Jutta Gampe); and Mercyhurst University (Stephen D. Ousley). Although each researcher specializes on a different component of the work—the bones, the math, or the software—the individual and collaborative projects, which include this dissertation, each contribute important components to the larger research agenda.

In 2014, this dissertation and a more extensive research team project were simultaneously funded. These projects are the first full-scale attempts to develop a statistically valid method of estimating age with low and known error across the entire adult lifespan using traits from many locations throughout the skeleton. Both projects focus on expanding the existing TA framework to include features throughout the skeleton that provide information over the entire adult lifespan (15–105 years). The research team project, funded by the National Institute of Justice (NIJ) (PI: Milner; Co-PIs: Boldsen and Ousley), focuses on the development of a method for forensic applications using skeletal samples from modern populations of differing ancestry on four continents. This dissertation, funded by a National Science Foundation Dissertation Research Improvement Grant (DDRIG #1455810; GR Milner, Advisor), focuses on many of the same skeletal traits, using a similar statistical procedure, but investigates the applicability of this age-estimation approach to a single broadly defined ancestry group over time. If the new TA method consistently produces improved adult age estimates for modern individuals from populations around the world and for those born over hundreds of years, we will have, for the first time, a method that can be confidently applied in both forensic and archaeological settings.

This dissertation is divided into four phases: 1) new trait identification and refinement using known-age reference collections; 2) collaborative trait refinement, error testing, and method development with the research team; 3) validation study, using both modern and historical skeletons; and 4) archaeological application. Phases 1, 3, and 4 are the primary components of this dissertation and were funded by two Penn State Department of Anthropology Hill Fellowship Grants and NSF. Phase 2 contributes substantially to subsequent phases of the dissertation, but is a part of the collaborative NIJ-funded research project. In Phase 1, individuals from four collections in North America are used to select features to be incorporated into a revised TA procedure. In Phase 2, these features are refined by the research team and combined using a simplified version of the existing TA procedure. Variations of this method are then tested on two temporally and geographically distinct known-age European skeletal samples in Phase 3, and applied to two Danish archaeological samples in Phase 4. The validation study demonstrates the accuracy of the new procedure and reveals the likelihood of individuals over 50 being detected in archaeological samples, if they indeed existed. The results of these tests and applications also indicate the extent to which well-documented errors associated with adult age estimation potentially influence our understanding of population dynamics in the past. Even if few individuals over 50 are detected in the archaeological samples, improved age estimates will allow for more detailed study of key anthropological issues, such as variations in mortality in different cultural and environmental settings and the evolutionary relationship between human menopause and longevity, by other researchers.

Phase 1: New trait identification and refinement using known-age North American reference collections

The approach used to estimate adult age by most standard methods is straightforward. Skeletal variation is described for one anatomical area, such as the pubic symphysis or auricular surface, using a fixed set of phases that each include features assumed to change together throughout life. Each phase is associated with a point estimate and age range calculated directly from its distribution in a skeletal reference sample. Unfortunately, while many features described in such methods experience age-related changes, at least for parts of the lifespan, there is great variation in how they do so relative to other traits within the same individual (Algee-Hewitt, Tersigni-Tarrant, & Shirley, 2013; Purves, Woodley, & Hackman, 2011). Additionally, the known-age reference collections where these features are documented are typically small and, more importantly, biased samples of the living populations that they represent (Bocquet-Appel & Masset, 1982; Usher, 2002). As a result, the age distributions of traits in skeletal samples do not necessarily reflect the true age distributions of the traits in living populations. For these reasons, traditional age-estimation methods have proved to be an inadequate way to capture complex biological variation.

Unpublished work by members of the research team has demonstrated that, after training on hundreds of known-age skeletons, experienced observers can produce age estimates with higher accuracy and finer precision than any existing method (Milner & Boldsen, 2012c). This indicates that significantly more age-related variation exists in the skeleton than is captured by the relatively few anatomical structures used in standard techniques. The process of deconstructing the features used to make experience-based age assessments began in conjunction with the development of the TA procedure in the late 1990s (Boldsen, Milner, Konigsberg, & Wood, 2002). Almost a decade later, this list was expanded by George Milner based on extensive skeletal experience and input from other research team members. Preliminary tests were conducted by Milner with several traits at the William M. Bass Collection, and by other members of the research team who used these and other traits and looked at additional skeletons. Beginning with a list of features from Milner's initial work, the goal of Phase

1 is to identify and define a set of macroscopic features throughout the skeleton that collectively provide age information for the entire adult lifespan. Phase 1 is divided into two subphases—1a) preliminary trait identification and refinment and 1b) primary trait selection.

In Phase 1a, four rounds of data collection, with samples of 100 to 200 skeletons from the Bass Collection in Knoxville, TN, were used in an interative process of trait identification, data collection, analysis, and refinement to select traits with promising age-related patterns. Although the initial list of features was primarily developed by Milner, additional features were suggested by Getz for each round of data collection. The development and refinement of trait definitions, as well as the evalution of each set of preliminary data, were highly collaborative endeavors between Milner and Getz. In Phase 1b, 53 traits identified in preliminary tests were investigated using 1010 individuals from four skeletal collections in the United States and Canada—the Bass, Maxwell Museum Donated, UI-Stanford, and JCB Grant collections. These data were used to select traits for further investigation in Phase 2.

Phase 2: Collaborative trait refinement, error testing, and method development

Phase 1 uses extensive work by an informal research team as a starting point to identify and refine a suite of new age-informative skeletal features using multiple rounds of preliminary data collection and the analysis of a large primary reference data set. However, for these traits to be of use to the osteological community, they must be formally incorporated into an easy-to-apply age-estimation procedure and disseminated in an accessible format. These needs are addressed as core components of the collaborative NIJ-funded research team project. Using modern skeletal reference collections from four continents, the goals of the NIJ project are to: 1) define new age-informative skeletal features; 2) assess inter- and intra-observer error; 3) establish trait age distributions in populations of diverse ancestry; 4) investigate theoretical and statistical improvements to the existing TA framework; and 5) develop materials for the dissemination of the new method to the osteological community, including an illustrated scoring manual and a user-friendly computer program.

Portions of the NIJ project contribute to this dissertaiton as Phase 2. Prior to the start of NIJ-funded data collection, the traits identified in Phase 1 were combined with additional traits suggested by other members of the research team. The features were then assessed in modern skeletal samples representing populations from North America, Europe, Africa, and Asia. After each round of data collection and analysis, existing trait definitions were refined by experienced osteologists, both native and non-native English speakers, with various levels of familiarity with the features. Definitions were reformatted, diagrams were added, and trait variants and scoring exceptions were extensively documented through photographs to create a scoring system that is easy to teach and apply with low error. Features demonstrating consistant age-related patterns across populations were selected for use in the new TA method. As a research assistant on this project, I worked closely with Milner and Boldsen to plan data collection trips, create forms, select samples, collect skeletal data in the field, photographically document skeletal traits, enter and analyze data, manage the content of the trait scoring manual, and substantively contribute to co-authored posters and presentations.

In this dissertation, a simplied version of the TA statistical framework developed by the research team is used to analyze the primarily reference data collected in Phase 1. In Phase 3, variations of the new TA procedure are tested using combinations of statistical models, reference samples, prediction intervals, and numbers of traits to identify the strengths and weaknesses introduced by each approach. The results of the new TA method applied single ancestry group

over time provide a foundation for more extensive comparative work by the research team using our large data set from modern populations around the world.

At the time of this writing, the NIJ project is entering its third, and final, year. At the completion of the project, anticipated for the end of 2017, data collected by the team and a portion of the Phase 1 and Phase 3 data from this dissertation will be incorportated into an updated version of the TA program (Boldsen et al. 2002). This program will allow age information from a large number of selected features to be combined with information about population structure in order to produce accurate and precise age estimates for the entirety of adulthood. At this time, all of the reference sample data have been collected and analyzed. The team is currently working on specific aspects of method development, including approaches for mitigating the effects of trait correlations and the use of population prior distributions. The final new TA method is anticipated to be released in 2018 as both a stand-alone program, as well as part of the commonly used software program, Fordisc (Ousley & Jantz, 2015).

Phase 3: Validation study

In Phase 2, procedures developed by the research team are used to fit variations of several statistical models to the primary reference data collected in Phase 1. Using these models, the probability of each feature existing over the adult lifespan is estimated for the entire combined sample, for males and females separately, and for each of two death cohorts. In Phase 3, these data for each trait are used in various combinations to produce age-at-death estimates for two temporally and geographically distinct European samples—individuals born in the nineteenth and twentieth centuries in Athens, Greece (Athens Collection), and in the seventeeth, eighteenth, and nineteenth centuries in London, England (St. Bride's Crypt). The two test samples represent populations spanning more than 300 years that varied significantly in their nutrition, disease exposure, and access to medical care, both from each other and from the Phase 1 reference collections.

Age estimates produced using variations of the new TA method are compared to estimates generated by commonly used age-estimation methods and the original TA program, which uses features only from the cranium and pelvis. These estimates are also compared with experience-based estimates to show the degree to which the new TA method captures the information throughout the skeleton used by experienced osteologists in expert age assessments. The performance of the method on the Athens skeletons demonstrates the method's applicability in modern, forensic contexts, while those from St. Bride's crypt are an appropriate comparative sample for the Danish archaeological skeletons evaluated in Phase 4.

Phase 4: Archaeological application

In Phase 4, the newly validated TA procedure is applied to two contextually well-documented samples from the Danish city of Horsens—Ole Worms Gade (ca. 1100–1500 CE) and Klosterkirke (ca. 1500–1800 CE). These samples represent populations from the Middle Ages and Early Modern period, in which, based on trends seen in historical documentation, we would expect to see differences in mortality profiles. These samples are evaluated using several standard age-estimation methods, the original TA method, and the new TA procedure. Mortality profiles are derived for each of the samples using the data from each method. These age-at-death distributions are compared with one another and to those produced for the known-age skeletal samples evaluated in Phase 3 to assess the potential biases introduced into paleodemographic analyses by each method.

Traditional procedures produce age-at-death distributions with method-specific charateristics that do not resemble one another or, based on the Phase 3 tests with known-age samples, reality. Age distributions produced using data from the new TA method reveal subtleties in population age structure, including increased numbers of individuals in older age categories, which are not apparent using data from traditional methods.

Issues of sample formation, age estimation error, and population structure are considered in the discussion of how well the improved age estimates bridge the gap between archaeological data and historical sources. This application demonstrates the high degree to which the use of different age-estimation techniques likely impacts the conclusions drawn from archaeological data and supports the necessity of re-evaluating previously investigated samples.

Dissertation Structure

Chapter 2 briefly introduces fundamental concepts related to human senescence, skeletal aging, and the use of skeletal data in archaeological and paleodemographic contexts. This chapter discusses theoretical and practical issues surrounding the process of human aging and provides an overview of adult skeletal age estimation. It includes a brief history of the field, discussion of what has been learned from nearly a century of investigation with skeletal collections around the world, and an overview of key issues in the use of skeletal age estimates in paleodemographic studies. Chapter 3 describes Phase 1—the process of age-informative trait identification, refinement, and selection from a large reference sample of known-age North American skeletons. Chapter 4 summarizes key components of the NIJ-funded research project currently underway by an international team as they apply to this dissertation. Chapter 5 presents the results of the validation study with modern individuals from Athens, Greece, and historical individuals from St. Bride's crypt in London, Engand. The performance of the new skeletal features is compared with experience-based age estimates, original TA using the cranium and pelvis, and other traditional age-estimation methods. The implications of differences among these methods for archaeological applications and the potential for future improvements in adult age estimation are discussed. Chapter 6 presents the results of various age-estimation methods applied to two Danish archaeological samples and discusses the implications of these results for other archaeological investigations. Chapter 7 summarizes the lessons learned during the investigation of traits, results from the known-age tests and archaeological application of the new procedure, and limitations of this work. The implications of these results for forensic and archaeological applications and avenues of future research are also discussed.

CHAPTER 2: AGING OF THE ADULT SKELETON—THEORY & PRACTICE

For well over 100 years, anatomists and physical anthropologists have been interested in agerelated variation in the human skeleton (e.g., Davies, 1913; Dwight, 1890a, 1890b; Parsons & Box, 1905). Fundamental to these investigations is that bone is a living, dynamic tissue capable of adapting and repairing itself in response to activity, disease, and the demands of life (M. Cox, 2000; Seeman, 2008). From the beginning of skeletal formation *in utero* until well into the third decade of life, the growth and development of bones and teeth can be used to estimate age. These estimates are inherently more precise than those based on "degenerative" changes in the adult skeleton, which begin almost as soon as skeletal development is complete and are influenced by a multitude of poorly understood factors (M. Cox, 2000; Franklin, 2010; İşcan & Loth, 1989; Mays, 2015). Although methods exist to adequately estimate age-at-death from the skeletal remains of infants, children, and adolescents, methods for estimating age throughout adulthood with a useful degree of precision remain elusive. Estimating age for individuals over the age of 50 years is particularly problematic (Falys & Lewis, 2011; Milner & Boldsen, 2012b).

Although it is ultimately chronological age—the amount of time that has passed since birth—that is most important in forensic, archaeological, and paleodemographic investigations, this information must be estimated from features of the skeleton (Milner & Boldsen, 2012b). The development and application of skeletal age-estimation methods relies on several critical assumptions: first, that skeletal variation in the adult skeleton can be identified and classified in ways that can be accurately and reliably applied by others (Milner & Boldsen, 2012b); second, that the traits, and the scoring schemes developed for them, correlate with the passing of time in a predictable and meaningful way (M. Cox, 2000); finally, that the features change in the same way and at the same rate in all populations, or at least readily identifiable groups of them, such as males and females (Milner & Boldsen, 2012b).

It is a commonly believed among both forensic anthropologists and archaeologists that agerelated skeletal changes occur at different rates both within and between the sexes and among populations; therefore, population and sex-specific standards will provide better age estimates for adults (M. Cox, 2000; Eliopoulos, Lagia, & Manolis, 2007; İşcan & Loth, 1989; A. Schmitt, Murail, Cunha, & Rougé, 2002). Although females are typically less well represented than males in skeletal reference collections, obtaining reasonable numbers of males and females is still possible by using multiple samples. Unfortunately, the number of well-documented skeletal reference collections in the world is far fewer than the number of populations that forensic anthropologists are interested in. The reference collections that do exist are typically relatively small, and inherently biased, samples of the populations they supposedly represent (Hunt & Albanese, 2005; Komar & Grivas, 2008; Usher, 2002). Therefore, population-specific standards are based on relatively small numbers of individuals who may or may not be a good match to the population or the individual being analyzed (Milner & Boldsen, 2012b). One approach to this problem is to use a large and diverse reference sample that is likely to capture the range of morphological variation that could conceivably be expected in all human populations. Although merging data from different groups is likely to decrease the precision of age estimates (i.e., broaden them), this conservative approach is ideal for many forensic and archaeological situations where the factors affecting skeletal variation are largely unknown and a close match to a reference sample is not possible (Milner & Boldsen, 2012b; A. Schmitt et al., 2002; Townsend & Hammel, 1990).

Aging and the Adult Skeleton

Our ancestors have likely been aware of the concept of aging—the predictable degeneration of the body with time, eventually ending in death—for at least the last 50,000 years (Arking, 2006b). However, it was not until the late nineteenth century when the first modern scientific inquiries related to the human lifespan and its relationship to social and environmental variation began to take shape (Arking, 2006b; Weismann, 1891). In the century that followed, continuing until today, investigations into the biology of aging became fundamental to the study of being human (Arking, 2006b).

As the study of human aging has expanded in both depth and breadth, the concept has been variably defined for use in a multitude of contexts (Arking, 2006b). In this dissertation, the most useful definition of aging makes a broad distinction between development and essentially everything else. Development consists of early processes that enhance functional capabilities, whereas aging consists of all subsequent changes that either diminish or have no effects on the ability to function (Kohn, 1978). Importantly, aging increases the probability that an individual will die over time and decreases their ability to withstand extrinsic stresses (Arking, 2006b). It is this relationship between the effects of aging and factors such as climate, food availability, diet, and disease exposure that fundamentally determines much of the demographic structure of a population (Weiss, 1981).

Chronological versus biological age

Physiological aging, including changes in the skeleton, is not a simple process that perfectly correlates with the passage of time (Arking, 2006b). In most of the modern world, it is common for an individual's chronological age to be precisely documented. Chronological age is linear and occurs at the same rate in all individuals and populations where time is measured in conventional units, and it is this metric that is of most value to forensic anthropologists and archaeologists. For example, chronological age is key to identifying individuals in forensic settings, estimating the age structure of populations in paleodemography, and calculating the number of person-years of risk in paleoepidemiology. Unfortunately, the link between observable biological changes in the adult skeleton and chronological age is poorly understood and influenced by a number of complex interrelated factors (M. Cox, 2000; Mays, 2015).

An array of possibilities, including genetics, endocrine function, disease, joint biomechanics, diet, body size, body composition, activity level, and even climate, among others, have been invoked to explain perceived differences in the timing and appearance of age-related features among individuals and samples (Belkin, Livshits, Otremski, & Kobyliansky, 1998; İşcan & Loth, 1989; Mays, 2015; Merritt, 2015; Wescott & Drew, 2015). Disappointingly, in the best-case scenario, investigations of skeletons can provide only very general statements about the correlation of features with age based on small and often biased samples of living populations. While these investigations provide critical starting points for future investigations of factors influencing the aging process, they cannot offer convincing evidence for causal relationships between skeletal traits and social conditions, biology, or the environment (Arking, 2006b). Although much of the variation among individuals may have a biological explanation, many of the changes within each individual may simply be the result of random factors (Weiss, 1981). Large samples, analyzed with appropriate statistic models, are necessary to identify patterns in the aging process and to understand the relationship between variation in aging and the factors that influence it (M. Cox, 2000; Weiss, 1981).

Activity, aging, and disease

In his first paper discussing age-related changes in the pubic symphysis, Todd (1920, p.287) writes that "many of these modifications which appear successively during adult age are on the border-line between the anatomical and the pathological." For the features throughout the skeleton used in this dissertation, the distinction between aging and disease is an important terminological issue. Anthropologists have historically made hazy, and sometimes arbitrary, distinctions between skeletal features related to "normal" aging, disease and activity depending on their educational background, experience, and research objectives. Although some early work recognized the complex interplay between disease processes and age (Sashin, 1930; T. D. Stewart, 1958), the majority of studies investigating musculoskeletal stress markers and degenerative joint disease have primarily focused on the relationship with habitual activity patterns (Larsen, 1997; Milella, Belcastro, Zollikofer, & Mariotti, 2012; Pearson & Buikstra, 2006). Recently, several studies of features traditionally believed to be primarily related to repetitive activity and disease, including osteoarthritis and ossification at fibrous and fibrocartilaginous attachment sites, found these features to be more highly correlated with age than other factors (A. F. Cardoso & Henderson, 2010; Milella et al., 2012). It is also often impossible, and perhaps even detrimental, to make a distinction between "disease" and "aging" because of the increased likelihood of one with the other. For example, although younger individuals may develop osteoarthritic changes in particular joints because of trauma, obesity, or extreme physical activity, the probability of this occurring in multiple areas of the skeleton increases with age, as do other risk factors including muscle weakness, obesity, and hormonal dysregulation (Felson et al., 2000).

Here Kohn (1978) offers another broadly useful definition: aging is considered a progressive and irreversible process that occurs in all members of a population and accelerates at maturity, whereas disease occurs in only a fraction of all individuals. With the exception of several developmental features, such as epiphyseal union of the clavicle, the features investigated in this dissertation begin to appear once the skeleton is fully formed and probabilistically increase in frequency with advancing chronological age. Although many of these features may have, at one time or another, been dismissed as markers of activity or disease by other researchers, no a priori assumptions were made in the initial selection of features for investigation.

Features were initially selected for investigation based on a subjective judgment of their age-informative potential and were ultimately selected for inclusion in the new TA procedure based on the appearance of consistent age-related patterns in multiple samples. In each individual skeleton, features were excluded only if they were clearly affected by trauma or a disease process, or were part of a skeletal abnormality. For example, in an individual affected by cancer, features in direct contact with an abnormal growth or a resorptive area were not scored. Features elsewhere in the skeleton, however, were included despite the possible systemic effects of the disease. This decision is consistent with the goal of creating a broadly applicable reference sample that captures as much of the morphological variation within a population as possible.³

_

³ This scoring procedure also makes sense for archaeological or forensic applications where often only a portion of the skeleton is present. For example, if only the thorax and upper limb of an individual are recovered, cancer that was active only in the pelvis could not be detected and the features in the available elements would be used to estimate age. The reference samples used in this dissertation and the NIJ-funded project (described in Chapters 3 and 4) capture the variation potentially introduced by disease processes.

Decline in the rate of increase of age-specific mortality rates

Several modern demographic studies have provided evidence for a decline in the rate of increase in age-specific death rates after the age of 80, particularly in females (Arking, 2006a; Horiuchi & Wilmoth, 1998). Two primary hypotheses have been proposed to explain this slowing in late-life age-specific morality—the heterogeneity and the individual-risk hypotheses (Horiuchi & Wilmoth, 1998; Khazaeli, Xiu, & Curtsinger, 1995; Vaupel, Manton, & Stallard, 1979). The heterogeneity hypothesis speculates that the deceleration is the result of frailer individuals dying at younger ages, so that by the older ages, the survivors tend to be healthier and age more slowly (Horiuchi & Wilmoth, 1998). The alternative—the individual-risk hypothesis—has several variations that focus on possible causes of slowing mortality at older ages at the individual, rather than the population, level (Horiuchi & Wilmoth, 1998).

Most osteologists assume that individual variation in the skeleton accumulates with time so that by around 50 or 60 years of age it becomes so great that nothing meaningful can be said about age (Milner & Boldsen, 2012a). This view has led to the acceptance of methods of adult ageestimation that use extremely wide age ranges or terminal open-ended intervals (e.g., 50+ years) for the latter half of the lifespan (Milner & Boldsen, 2012a). Encouragingly, however, recently published work using existing Transition Analysis (TA), shows that this commonly held assumption might be unjustified (Milner & Boldsen, 2012c). Although skeletal variation increases between 45 and 75 years of age, it appears to decrease after this time, as reflected in a narrowing of predicted age intervals (Milner & Boldsen, 2012c). Furthermore, the same study and research conducted as part of this dissertation demonstrates that experienced observers can accurately predict age to the maximum extremes of the adult lifespan (Milner & Boldsen, 2012c). Selective mortality—the tendency for the frailest individuals at each age to be pulled out of the population—may be responsible for this effect (Milner & Boldsen, 2012c). In essence, selective mortality tends to reduce the amount of variation present at each age, so individuals who make it to the oldest ages, no matter how decrepit their skeletons may appear, tend to appear more similar than would any two individuals in middle age.

Traditional Approaches to Adult Skeletal Age-Estimation

The most common approach for creating an adult skeletal age-estimation method is relatively straightforward and has remained largely unchanged since its beginnings almost a century ago. First, an area of the skeletal is chosen and age-related trait variants are identified. The variants are either scored individually or grouped into collections of features, usually called phases or stages. A collection of documented skeletons of known age, sex, and ancestry (i.e., a reference sample) is then selected. Each individual is evaluated and classified as one of the variants or stages. A point age and range are calculated for each variant based on all individuals who were given each score, usually with the assumption that the ages of the individuals assigned to each variant have a normal distribution. Most commonly, the point age for each category is the mean age of all individuals assigned that score, while the range is plus and minus two standard deviations (i.e., an approximate 95% confidence interval). The feature variants are documented through written descriptions, photographs, diagrams, and, sometimes, casts of exemplars.

One of the most commonly used methods is the six-phase Brooks and Suchey (1990) pubic symphysis technique. Each phase is described as having a unique combination of features associated with a mean and range (Figure 2.0). The method was published with photographs and descriptions, and casts of exemplars for teaching and reference can be purchased. While these materials are extremely helpful, the examples were selected to represent ideal cases

where all features described for a phase are present and relatively easy to detect. Although groups of features are presented as moving together in lock step throughout life, in reality variation from the classic example is the norm, not the exception (Algee-Hewitt et al., 2013; Boldsen et al., 2002; İşcan & Loth, 1989; Purves et al., 2011).

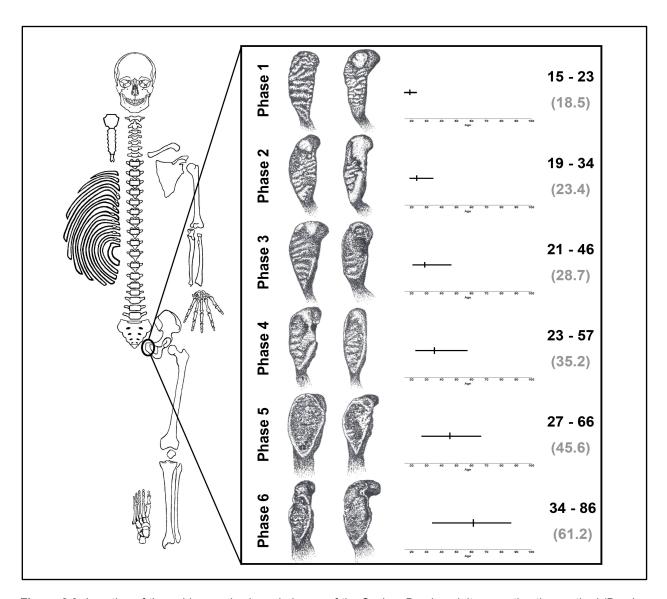


Figure 2.0. Location of the pubic symphysis and phases of the Suchey-Brooks adult age-estimation method (Brooks & Suchey, 1990). Early and late examples of each phase are shown, along with the associated mean (gray) and range (range) for each. Example figures drawn by George Milner from Tocheri et al. (n.d.) and Diane France casts.

This is particularly problematic because all individuals in each phase are assigned the same point age and range, regardless of how well the morphology is actually captured by the phase description. An implicit assumption of this approach is that all individuals assigned the same score are equally likely to have a chronological age that falls into the associated age interval. Thus, there is no way to express uncertainty in an age estimate other than selecting more than one category, which only further widens already impractical ranges.

Aggregating traits into multidimensional phases results in methods that fail to capture much of the age-related variation in complex anatomical units in a biologically meaningful way, increase error between observers, and ignore significant information about inter- and intra-individual skeletal variation (Algee-Hewitt et al., 2013). An alternative way to describe anatomic variation is the component approach, where individual features or areas of a bone are described using sequences of ordinal variants. The classic example of a component-based method is McKern and Stewart's (1957) work with the pubic symphysis, where five developmental stages for each of three components were defined. Unlike phases where features are assumed to occur together, each component is scored independently without considering other features. Some feature combinations were not observed in the original sample and were thought to be unlikely to occur, but point estimates and ranges for all possible sets of summed scores are provided. The component approach allows for a finer categorization of the age-related changes within a complex anatomical feature. However, the use of simplistic summary statistics, such as means and standard deviations, to produce fixed sets of possible age estimates for the summed component scores remains problematic.

To address this issue, TA was developed to take the component approach a step further. In this method, features of the cranial sutures, pubic symphyses, and iliac auricular surfaces are separated into independently scored components, each with a unidirectional series of variants. A form of logistic regression known as the cumulative logit is used to model the transition between ordinal feature stages in large samples of skeletons (Boldsen et al., 2002). The probabilistic age information for the traits and their character states in each skeleton is combined to produce a maximum likelihood estimate of age, which is then adjusted based on prior information about the age structure of the population the individual likely came from. This approach allows the age-related variation present in each feature to be combined in a statistically valid way. The result is a probabilistically tailored maximum likelihood point estimate of age with a confidence interval for each individual based on the skeletal features present and the consistency of the age information provided by each trait.

Although TA is a statistical step forward, the estimates produced still fall short of what is needed for forensic and archaeological applications, particularly for individuals between 45 and 75 years of age (Milner & Boldsen, 2012c). Two relatively recent paleodemographic applications of the method to archaeological samples from Postclassic Cholula and Contact-Period Xochimilco Mexico (Bullock, Márquez, Hernández, & Ruíz, 2013) and Late Classic Copan in Honduras (Storey, 2007) came to similar conclusions. Mortality profiles produced using TA estimates are more similar to what would be expected for ancient populations based on existing historical and ethnographic data than those produced using traditional techniques; however, greater precision for individuals over 50 is necessary for more detailed analyses.

Other researchers have also recently demonstrated that the application of new scoring systems or more advanced statistical techniques to features of the pubic symphysis and sacroiliac joint fail to greatly improve their ability to predict age (M. F. Anderson et al., 2010; Buk et al., 2012; Hens & Godde, 2016; Martins, Oliveira, & Schmitt, 2012). Using only these parts of the skeleton in any analytic approach seems to result in wide intervals and biased estimates of age for at least some portions of the adult lifespan. Combining information from more than one part of the skeleton is a commonly advocated approach for resolving these issues (Angel, Suchey, Iscan, & Zimmerman, 1986; Brooks & Suchey, 1990; DiGangi, Bethard, Kimmerle, & Konigsberg, 2009; İşcan & Loth, 1989; Lovejoy, Meindl, Mensforth, & Barton, 1985; McKern & Stewart, 1957; Meindl & Lovejoy, 1985; Saunders, Fitzgerald, Rogers, Dudar, & McKillop, 1992; Todd, 1920). Unfortunately, age estimates from standard methods (e.g., Brooks & Suchey, 1990; Lovejoy, Meindl, Pryzbeck, & Mensforth, 1985; Meindl & Lovejoy,

1985) are difficult to combine in a statistically meaningful way (Buckberry, 2015). Tests of methods that combine multiple features, such as the Complex and Multifactorial methods, indicate that while they produce some improvement relative to methods using only one part of the skeleton, they do not solve the issue of age-estimation bias in the latter half of the adult lifespan (Baccino, Ubelaker, Hayek, & Zerilli, 1999; Bedford et al., 1993; Lovejoy, Meindl, Mensforth, et al., 1985; Milner, Wood, & Boldsen, 2008; Saunders et al., 1992).

In sum, the systematic age-estimation bias resulting from the failure of standard features to keep pace with chronological age from middle age onward cannot, in isolation, be overcome by any statistical technique (Boldsen et al., 2002; Milner & Boldsen, 2012c). Furthermore, the problems associated with all existing adult age-estimation methods cannot be solved simply by combining features or methods that individually suffer from the same sort of issues (Milner & Boldsen, 2012a). To move towards more accurate and precise adult age estimates, statistical improvements must be paired with a broader array of age-informative traits.

Although every bone has the potential to show visible signs of aging (İşcan & Loth, 1989) only a limited fraction of the adult skeleton has been extensively investigated for age-related variation. By far, the most commonly used features to estimate age are from joints where there is no or limited movement, including the cranial sutures, sacroiliac joints, pubic symphyses, and sternal rib ends (M. Cox, 2000). These joints show many changes that can be categorized in various ways, but why these changes appear and the factors that influence them are not well-understood (M. Cox, 2000). It is also not uncommon to observe different levels of age-related change among various parts of the same individual (M. Cox, 2000; Ubelaker, 2000). Weiss (1981) cautions that while the systematic and predictable deterioration of the body is an immutable feature of our species, the specific pattern found in each individual can only be predicted in terms of vague probabilities. This sentiment has been echoed by physical anthropologists, who have noted that the overall pattern of skeletal aging can be highly variable from one individual to the next (Işcan, 1989).

Some additional features, such as vertebral osteophytes and changes in multiple portions of the scapula, have been investigated; however, they are typically viewed as containing insufficient information to adequately predict age (Graves, 1922; T. D. Stewart, 1958) because they provide only very general signals of "young" or "old" and are individually of little value (Milner & Boldsen, 2012c). With the continued publication of validation studies confirming the issues with commonly used features, a number of researchers have investigated new skeletal characteristics and revisited some that were previously dismissed (e.g.,DiGangi et al., 2009; Falys & Prangle, 2015; Listi, 2016; Listi & Manhein, 2012; Rissech, Estabrook, Cunha, & Malgosa, 2006). Disappointingly, the consensus remains that even using new scoring techniques and statistical approaches, these features individually contain insufficient agerelated variation to produce practically useful estimates.

Adult Age Estimates in Paleodemography

Paleodemographic analyses have three broad analytical steps that require assumptions to be made about the sample, the data, and the methods of analysis. First, skeletal data are collected using standard techniques, including those for estimating age. Second, these data are typically summarized in tables or charts or incorporated into more sophisticated procedures to produce quantitative descriptions for a group or population. Lastly, these results are interpreted to generate hypotheses or support conclusions about the relationship between population structure and changes in factors such as social conditions, biology, or environment. To use skeletal age-estimation methods to investigate past populations requires the

assumption of uniformitarianism (Howell, 1976). This concept, originally borrowed from geology, states that the forces acting the human species, and our biological responses to them, are the same now as they were in the past (Howell, 1976; Simpson, 1949). For adult age estimation, this assumption operates at both the level of the traits being assessed for each skeleton and the population-level patterns that collectively emerge from combining all of the individual estimates. An assumption must be made that aging occurred in the same way in the population used to develop the age-estimation method as it did in the population being investigated (Howell, 1976). In other words, the relationship between observable skeletal traits and chronological age must be assumed to the same in the more modern reference samples as it was in the past. Additionally, critical features that shape demographic patterns, such as greater susceptibility of infants and children to disease, maximum fertility levels, and maximum human lifespan, must be assumed to have not changed greatly over time (Howell, 1976; Milner et al., 2008). This is not to say that the number of children produced or the average lifespan does not vary over time or across populations, only that the maximum limits on these features are biologically determined and do not fluctuate significantly (Howell, 1976).

If the data obtained from an age-estimation method is vastly different than what was expected based on uniformitarian assumptions, there are two possibilities (Howell, 1982). The first is that the population under investigation is different from other similar groups for genetic, biological, or social reasons. This result is only possible if the sample adequately reflects the structure of the original population and the method being used produces accurate and unbiased estimates of age. The second is that the differences are merely the result of biases introduced by the sample analyzed or the particular age-estimation method used (Howell, 1982).

Skeletal samples

Archaeological skeletal assemblages are not representative samples of once-living populations. This is because at every step in their formation—death, deposition, preservation, recovery, and analysis—there are multiple factors that systematically bias which individuals are ultimately included in the final sample.

Archaeologists must first make the assumption that the individuals contributing to the mortality sample came from a single population that maintained a similar culture over time and had a constant level of fertility and mortality⁴ (Milner et al., 2008). However, the possibility exists that the living "population" contributing to the mortality sample was also changing, either through fluctuations in fertility, mortality, and migration, or potentially from a new group or series of groups moving into or through the area. The size and the demographic composition of the population can change over time, together or independently, and both affect the number of individuals entering the death sample during any period (Chamberlain, 2006). Careful excavation and analysis of the context of the remains is necessary to determine whether the mortality sample likely represents a meaningful unit of some population (Milner et al., 2008).

⁴ A population can be defined in biological or sociocultural terms. Elsewhere in this dissertation, the term "population" is used to broadly to refer to a group of individuals who had similar ancestry and lived in the same geographic area or time period. For example, males and females of European ancestry that lived the majority of their lives in post-WWII America are considered a population that is represented by the Bass Collection. Similarly, the individuals from the two archaeological sites in the Danish city of Horsens could be considered two separate populations because they are separated by several hundred years, or a single pre-industrial population from the same city, depending on the research question.

For the individuals in the mortality sample, in the words of the well-known article The Osteological Paradox, "There is one, and perhaps only one, irrefutable fact about the cases making up a skeletal series: they are dead. We never have a sample of all of the individuals who were at risk of disease or death at a given age, but only of those who did in fact die at that age" (J. W. Wood et al., 1992, p. 344). In any population at a specific time, the likelihood of death for individuals of different ages across the adult lifespan, or even among individuals of the same age, is not uniform. Heterogeneity in frailty—the tendency for some individuals at each age to be more susceptible to death and disease—is acted on by the external factors to produce selective mortality (J. W. Wood et al., 1992). So even individuals of exactly the same chronological age differ in their risk of death because of differences in sex, occupation, health and other factors. Even if, by chance, the individuals who happened to die are numerically representative of the age and sex composition of the population, those individuals would not have all of the same characteristics of the still-living individuals because they are the ones who did, in fact, die (J. W. Wood et al., 1992).

Although an entire cemetery may contain representatives of all segments of a population, additional bias is introduced when only a portion of the area is excavated. Practical considerations, including time, money, personnel, and equipment, are often the primary constraints influencing the extent of an excavation (D. H. Ubelaker, 1995). Within the area that is excavated, cultural and taphonomic factors can further influence how much of the population can potentially be represented and which of these individuals have critical skeletal elements suitably preserved for analysis (Konigsberg & Frankenberg, 1994; Nawrocki, 1995; D. H. Ubelaker, 1995). The presence of clothing or a shroud, depth of burial, and use of a burial container and its type, such as a wooden or lead coffin, influence the probability of the skeletal elements remaining in close proximity to one other and their potential for preservation (Nawrocki, 1995). Other taphonomic factors, such as animal scavenging, human disturbance, soil type and pH, vegetation, groundwater, and drainage, also influence the preservation of the remains (D. H. Ubelaker, 1995). These factors can vary greatly within relatively small areas and differ among individuals because of variations in age, sex, body size, and disease status (Nawrocki, 1995).

Figure 2.1 summarizes these issues. A living population that is changing in composition over time contributes members non-uniformly to a death sample. These individuals are potentially interred in different ways (i.e., with or without clothes or coffins, in individuals graves or in groups, at different depths) and in different areas of the cemetery, perhaps separated by age, sex, or social status (Konigsberg & Frankenberg, 1994). Of the entire potential death sample, often only a fraction is excavated and, of this portion, only some of the individuals will have the necessary skeletal elements present or preserved for analysis. In archaeological settings, some or all of these factors will shape the sample that is ultimately analyzed (Milner et al., 2008).

From the above discussion, it should be clear why the structure of the mortality sample is not the same as that of the living population that it represents (Chamberlain, 2006). This means that whether the data adequately represent the sample is a fundamentally different problem than if the sample can be used to investigate the population. Although both issues are critical for producing accurate interpretations of the past, different sets of assumptions and statistical tools are needed to address each problem. Ultimately, if the input data are a biased and inadequate representation of the sample, the application of any statistical analysis or modeling

technique to generate information about the population will produce meaningless results.5

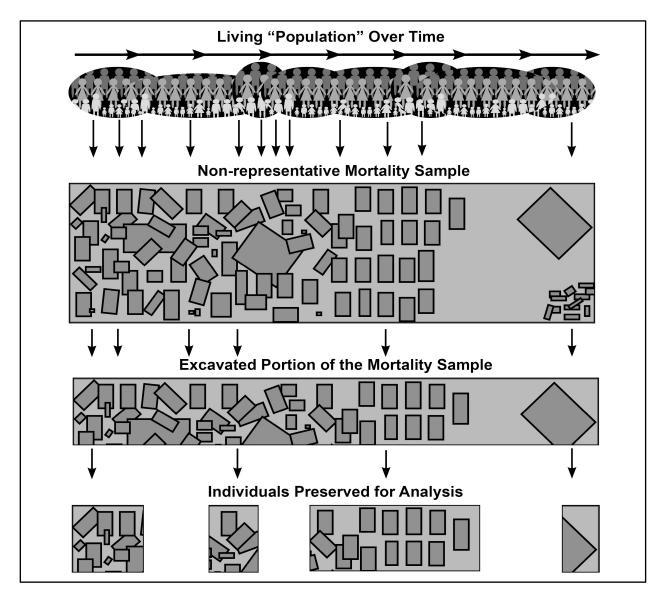


Figure 2.1. A schematic overview of the factors in archaeological investigations that may influence the skeletal sample that is ultimately available for analysis.

Applying age-estimation methods to skeletal samples

While the development of population-specific methods has become increasingly common in the past decade, these techniques are rarely used in most forensic and archaeological settings (Falys & Lewis, 2011; Garvin & Passalacqua, 2012). This is, in part, because interpreting the results of these methods is more complex when all other information available for a region or

⁵ There will always be some unknown, and unknowable, level of error in archaeological samples that can be mitigated through the use of various statistical and modeling procedures (Milner et al., 2008). It is the systematic and identifiable errors in existing procedures that should not be accepted without question.

population is based on the results of standard techniques (Falys & Lewis, 2011). Although the differences found between populations may be real, it is also possible that they are simply the result of errors introduced by the methods used, the structure of the test sample, or random variations in the particular individuals analyzed (Wärmländer & Sholts, 2011).

Paleodemographic results typically show that few individuals survived past 50 years of age, and large numbers died between their 20s and 40s. This pattern does not conform to model life tables based on plausible age-specific fertility and mortality values for modern and historical populations, including those documented ethnographically, such as the Aché, Gainj, !Kung, and Yanomamö (Coale & Demeny, 1983; Hawkes & Blurton Jones, 2005; Hill & Hurtado, 1996; Howell, 1979, 1982; Neel & Weiss, 1975; J. W. Wood & Smouse, 1982).

Although the source of these discrepancies is unclear, this pattern is found in essentially all archaeological populations, despite large cultural, temporal, and geographic differences among them (Chamberlain, 2006; M. Cox, 2000; Milner & Boldsen, 2012b). Furthermore, research on known-age skeletal collections from around the world demonstrates that methods consistently underestimate age for middle and old-age adults (e.g.,Gocha, Ingvoldstad, Kolatorowicz, Cosgriff-Hernandez, & Sciulli, 2015; Murray & Murray, 1991; Rissech, Wilson, Winburn, Turbón, & Steadman, 2012; Sakaue, 2006; A. Schmitt, 2004). Together, these data call into question whether published mortality profiles are an accurate representation of the prehistoric past, or if they are the product of age-estimation error—a debate with a 40-year history (Bocquet-Appel & Masset, 1982; Howell, 1982; Konigsberg & Frankenberg, 1994; Lovejoy et al., 1977).

Significance of accurate and precise age estimates

If appreciable numbers of individuals lived beyond 50 years of age in the past, the use of standard age-estimation methods results in a substantial loss of information for paleodemographic analyses. Because traditional methods often group all individuals in the latter half of the adult lifespan into an extremely wide or open-ended age category (e.g., 50+), they prohibit the examination of interesting questions related to variations in mortality in different cultural and environmental settings. Even if few individuals in the past lived beyond middle age, the known errors associated with adult age-estimation have potentially serious implications for our understanding of changes in age-specific mortality rates, social structure, and the capacity to mobilize labor for essential household and community tasks.

Additionally, although not addressed in this dissertation, there is considerable potential for improved age estimates to aid investigations of the timing and relationship between the evolution of human menopause and longevity (Gavrilova & Gavrilov, 2005; Lahdenperä, Lummaa, Helle, Tremblay, & Russell, 2004; Lahdenperä, Russell, & Lummaa, 2007; Lahdenperä, Russell, Tremblay, & Lummaa, 2011; Leonetti, Nath, Heman, & Neill, 2005; Peccei, 1995, 2005; Shanley & Kirkwood, 2001; Shanley, Sear, Mace, & Kirkwood, 2007). Adaptive hypotheses, including the mother (Peccei, 2001b; Williams, 1957), grandmother (Hawkes, O'Connell, Jones, Alvarez, & Charnov, 1998), and reproductive conflict hypotheses (Cant & Johnstone, 2008; Cant, Johnstone, & Russel, 2009), involve the fitness benefit resulting from the early cessation of personal reproduction and the redirection of energy into existing children or grandchildren. Non-adaptive hypotheses, such as the "new" grandmother (Peccei, 2001a) and male longevity hypotheses (Kaplan, Hill, Lancaster, & Hurtado, 2000; Marlowe, 2000; Tuliapurkar, Puleston, & Gurven, 2007), view menopause as a degenerative part of the human life course that we only now reach because of recent dramatic increases in longevity (Austad, 1994; Holman, O'Connor, & Wood, 2006). These hypotheses have been examined using historical data and modern hunter-gatherer and subsistence agricultural groups (Beise, 2005; Hawkes, O'Connell, & Blurton Jones, 1997; Hill & Hurtado, 1996; Lahdenperä et al., 2007; Leonetti et al., 2005; Shanley & Kirkwood, 2001; Shanley et al., 2007; Skjærvø & Røskaft, 2013). However, these populations are rare in comparison to skeletal collections, and all such groups are affected in some way by contact with the modern world, so there is great need to test these hypotheses using prehistoric populations that lived in social and physical environments more similar to those where these features evolved (Austad, 1994). To do so requires more accurate, precise, and unbiased age estimates than are possible using existing methods. For example, one critical assumption of the grandmother hypothesis is that females in the prehistoric past commonly lived long enough to experience menopause (Austad, 1994), which currently cannot be evaluated for most of human existence. This is also true for investigations of non-adaptive menopause hypotheses that assume that lengthening of the human adult lifespan has occurred over the past several centuries (Austad, 1994).

Moving Forward: This Dissertation & the NIJ Project

This dissertation is a complementary component of the NIJ-funded international research effort that is simultaneously grappling with the statistical, osteological, and technological challenges of the adult age-estimation problem (see Chapter 4). Both this dissertation and the larger project take the position advocated by one of the founders of forensic anthropology, T. Dale Stewart (1958, p.144) that "there seems little doubt that every part of the skeleton carries sign of ageing although as yet almost all of them remains to be full identified and analysed [sic] for practical application." Because different skeletal features are likely to reflect various aspects of the aging process depending on their position, structure, and function within the body (İşcan & Loth, 1989), using information from the entire skeleton, rather than a single region is likely to produce better results.

To improve adult age estimation, what is needed is a suite of skeletal characteristics that collectively demonstrate age-related variation for the entire adult lifespan, a more sophisticated way of capturing this information, and a method for producing individualized age estimates with associated measures of certainty. The statistical framework of the original TA method (Boldsen et al., 2002) derives age estimates from dozens of independently scored traits while addressing many of the statistical issues present in other methods. Thus, the extension of this procedure to features distributed throughout the skeleton, especially those providing information in middle to old age, is a logical next step in the quest to improve adult age estimates.

The first three phases of this dissertation focus on the identification and testing of skeletal features that provide relatively consistent age-related information in six diverse groups of individuals of European ancestry born over several hundred years. The age-estimation procedure developed in Phases 1 and 2 is investigated using both modern and historical skeletal samples in Phase 3. In Phase 4, the best-performing variation of the new TA age-estimation method, based only on the European samples evaluated in the previous phase, is applied to two Danish archaeological samples. Because the success of paleodemographic investigations ultimately depends on the accuracy of the input data used in the analysis, Phase 4 focuses on the performance of commonly used age-estimation methods, existing TA, and the new TA procedure. The implications for all bioarchaeological and paleodemographic studies that rely on skeletal data are discussed.

CHAPTER 3: PHASE 1—PRELIMINARY TRAIT IDENTIFICATION & REFINEMENT USING NORTH AMERICAN SAMPLES

Age-estimation papers for subadults tend to focus on overall patterns in the skeleton or dentition, while those for adults are heavily weighted towards the development, testing, and refinement of procedures based on relatively few features. The regions of the adult skeleton that receive the greatest attention are those that experience numerous changes throughout life, which can be divided or grouped in a variety of ways. The ongoing cycle of feature investigation, method creation, validation, and revision has resulted in many publications focusing on the cranium, pelvis and ribs. Appendix A is an extensive, but not exhaustive, list of over 200 publications related to the regions of the skeleton most commonly used for adult age estimation. The list includes general investigations of age-related skeletal variation, methods for the estimation of age, validation studies, and forensic and archaeological applications of specific techniques. Disappointingly, as studies investigating standard skeletal features have proliferated, so too has evidence for deficiencies in the existing techniques and the ways in which they are applied (Falys & Lewis, 2011; Garvin & Passalacqua, 2012).

The most notable weakness of standard methods is ubiquitous age-estimation bias. Existing methods tend to slightly overestimate age for young adults, but work reasonably well for individuals under around age 40. After this time, however, essentially all existing techniques underestimate age for the rest of the adult lifespan. This systematic age-estimation bias is the result of both statistical and biological issues in existing techniques. One long-recognized statistical culprit is that age intervals and point estimates, usually means or midpoints, are directly influenced by the age composition of the reference sample originally used to document the phases or individual features of each method (Bocquet-Appel & Masset, 1982; Buckberry, 2015; Konigsberg & Frankenberg, 1994). However, even if this statistical problem is eliminated, age-estimation bias will remain because the skeletal features used in existing techniques fail to keep pace with chronological age after around age 50 (Milner & Boldsen, 2012c; Saunders et al., 1992). Furthermore, what little age-related variation is present in the sutures, ribs, and pelvis is masked when age-informative features are combined into phases or summed scores. Aggregating traits into multidimensional stages also increases error between observers and ignores potentially significant inter- and intra-individual skeletal variation (Algee-Hewitt et al., 2013; Kimmerle, Prince, & Berg, 2008).

Impractically large age ranges, relatively low accuracy, and age-estimation bias make existing techniques inadequate for many medicolegal and archaeological applications. To resolve these issues, both the biological and statistical components of the age-estimation problem must be addressed. The goal of Phase 1 is to identify and define a set of age-related skeletal features throughout the skeleton that collectively provide information for the entire adult lifespan (15-105 years). Traits are first identified and refined in multiple rounds of preliminary investigation before a much larger, primary dataset is collected from North American and European individuals born in the nineteenth and twentieth centuries. Collaborative work with an international research team investigates and refines these features using populations from three additional continents as part of Phase 2. A quantitative framework for combining a subset of these features is tested in the validation study in Phase 3 and applied to archaeological samples in Phase 4.

⁶ A multitude of conference presentations, Master's theses, and PhD dissertations related adult skeletal age estimation were not included in this list. An exception was made, however, for two theses (Baker, 1984; Masset, 1982) that are commonly cited in relevant publications and reference manuals.

Identifying and Defining New Age-Informative Skeletal Features

Beginning in 1996, Dr. George Milner (Penn State) and Dr. Jesper Boldsen (University of Southern Denmark) began collaborative work on the adult skeletal age-estimation problem. A primary objective of their work was to address the systematic age-estimation bias that plagues all adult age-estimation methods. Because it was commonly believed that inaccuracy and ageestimation bias were primarily the result of statistical flaws and not the skeletal features used, the initial work focused on developing a new analytical approach for the pubic symphyses and auricular surfaces. The cranial sutures were also included despite their well-known limitations because isolated crania are often recovered in archaeological and forensic contexts. The three areas of the skeleton were separated into components and age information from each of the scores was combined using a new statistical framework that they called Transition Analysis (TA) (Boldsen et al., 2002). This technique represented a promising advancement. However, in the decade that followed, Milner recognized that osteologists with extensive skeletal experience were able to produce adult age estimates with greater accuracy and less bias than existing techniques, including TA. These experience-based estimates provided the first evidence that significantly more age-related variation exists in the adult skeleton than is being effectively captured by existing techniques. Because experience-based age assessments perform similarly well for the entire adult lifespan, deconstructing the components of these estimates is a logical first step towards producing unbiased age estimates for all of adulthood.

Starting with a list of potentially age-related features compiled during their initial work, Milner and Boldsen collaborated with members of their informal research team over the next decade to define over 75 traits that could possibly be contributing to experience-based estimates. These traits include, among others, the presence and development of exostoses, osteoarthritic changes in synovial joints, and modifications of bone texture and shape. Milner and other members of the research team, including Svenja Weise, Peter Tarp, and myself, independently carried out preliminary investigations and refinement of some traits in small forensic and archaeological samples as funding and time permitted. Some of the traits have previously been investigated by other researchers in small or regionally-specific samples (e.g, Albert & Maples, 1995; Black & Scheuer, 1996; Calce, 2012; DiGangi et al., 2009; Graves, 1922; Passalacqua, 2009; Rissech, Estabrook, Cunha, & Malgosa, 2007; Scoles, Salvagno, Villalba, & Riew, 1987; Snodgrass, 2004; Van der Merwe, Işcan, & L'Abbè, 2006). No study, however, has looked at such a wide variety of features, or has examined traits from many parts of the skeleton simultaneously.

This dissertation investigates the variation present in a wide array of features in a large, temporally diverse sample. This work is further developed and expanded to an even larger, geographically heterogeneous sample of modern populations in the collaborative NIJ-funded project (Co-PIs: Milner, Boldsen, and Ousley) where I am currently employed as a research assistant. Chapter 4 provides an overview of the NIJ project.

Using logistic regression to select new age-informative skeletal traits

Analyzing individual skeletal features or aggregated groups of them (i.e., phases) using simple statistical measures, such as means and standard deviations, results in large, overlapping age ranges (see Figure 2.0). Binary features—those scored as only present or absent—analyzed in the same way produce even wider age ranges that contribute little to the estimation of adult age. Additionally, simply calculating the mean age of individuals with and without a trait only provides information that is directly influenced by the age distribution of the reference sample.

If an alternative reference sample was selected containing a different proportion of young or old individuals, the point ages and ranges assigned to each of the traits would also differ. Although now a widely acknowledged fact, this issue and its implications for paleodemography caused significant controversy when first introduced by Bocquet-Appel and Masset (1982) in their paper "Farewell to Paleodemography." In their article, the age distributions of several samples used to develop commonly used methods were juxtaposed with the age distributions produced for other samples analyzed using each of the techniques; this direct comparison demonstrates the biases imposed by the reference sample on the age estimates produced.

One of these examples—the McKern and Stewart (1957) method—has become the classic example of reference sample mimicry. This component-based pubic symphysis age-estimation method was originally developed using a sample of deceased American servicemen from the Korean War (McKern & Stewart, 1957). As would be expected for a group selected from activeduty soldiers, the sample was heavily weighted towards young men between the ages of 18 and 40. As a result, the method works well for younger males, but it cannot be applied to older individuals without producing significantly biased results. This is simply because the age intervals for each summed component score are directly based on the distribution of the traits seen in the original reference sample, which did not include any individuals over 50 years. In other words, the point ages and ranges for each skeletal score are lower than they would be if the entire adult age range were represented in the reference sample. This is especially true of features, such as breakdown⁷, that most often occur in individuals who are middle-aged or older. If this method was used to estimate the ages of individuals whose true ages spanned all of adulthood, the resulting age distribution would be skewed to resemble the reference sample. The issues and implications of biases imposed by reference sample composition have also been discussed more generally by many authors (e.g., Buikstra & Konigsberg, 1985; Hoppa & Vaupel, 2002; Konigsberg & Frankenberg, 1992; Van Gerven & Armelagos, 1983).

One simple approach for dealing with reference sample mimicry is to estimate the probability of a trait existing in an individual of each age in a population rather than relying directly on the trait's distribution in a reference sample (Boldsen et al., 2002; Buckberry, 2015; Chamberlain, 2006). Although the process of data collection for this probabilistic approach and traditional methods is similar, the way in which these data are analyzed and the age estimates generated from them are fundamentally different. In both cases, a skeleton in a reference sample only provides a single datapoint—a trait, phase, or component score at their particular age-at-death. In each skeleton, we cannot know when the individual first developed those features or when, or if, the additional features would have appeared if the person had not died. For traditional methods, this means that the point age and range associated with each features are based directly on the distribution of ages associated with the trait scores of deceased individuals from the skeletal reference sample. As discussed above, this approach results in biased estimates of age if the age distribution of the reference sample does not match that of the target population, or if the sample size is small, which is often the case.

The alternative approach advocated here is to use logistic regression—a type of analysis that models the probability of an event occurring based on the values of an independent variable—to model a probability distribution for each trait. A logit or probit model fitted to a large sample of binary trait data from a known-age reference collection provides the probability of a feature being present or absent at every age across the adult lifespan. Rather than simply calculating

⁷ Breakdown is considered the terminal category of the metamorphosis of the ventral and dorsal margins of the pubic symphyseal face that is characterized by pitting, erosion, or irregular bone growth that destroys the regular raised rim of the previous phase. It is the final stage (5) of McKern and Stewart's (1957) pubic symphysis component III.

the average age of individuals with and without the trait, the model estimates the likelihood of an individual of every age in the population having the feature. The larger the reference sample, the more precisely the model can estimate the age at which a feature is most likely to develop. The age at which an individual is equally likely to have or not have developed a feature is commonly referred to as the age-of-transition (i.e., the point at which the feature likely "transitions" from being absent to present in a population based on the logistic model fitted to the data using maximum likelihood estimation). The model takes the form of an "s-shaped" or sigmoid curve whose slope characterizes the age information present in a feature—the steeper the curve, the more informative the feature is about age.

Probability, Odds, and the Link Function

To understand logistic regression as it is applied here, it is critical to understand the difference between probability and odds. A probability is calculated by dividing the outcomes of interest by all possible outcomes. For example, the probability (p) of rolling a one on a standard six-sided die is one out of six or (0.167). Although probability and odds are often used colloquially to mean the same thing, odds are actually a ratio of probabilities. Odds are calculated by dividing the probability of an event occurring (p) by the probability that it will not occur (1-p). In the die example, the probability of rolling a one is one out of six (0.167), while the probability of not rolling a one is five out of six (0.833). Thus, the odds are (0.167) divided by (0.833) or 0.2. The importance of this will be discussed shortly.

The goal of a logistic regression model is to be able to estimate the probability of an event occurring based on the value of an independent variable. In this case, the independent variable is chronological age and the dependent variable is a binary skeletal trait. As previously discussed, each skeleton can have only one observation for a particular trait. Therefore, the score for each skeleton in the sample can be thought of as a Bernoulli trial—a random experiment in which there are only two possible outcomes and the probability of success (p) is the same for each independent trial (Ross, 2010). When a series of Bernoulli trials are combined, their outcomes are distributed as a binomial random variable with the parameters representing the number of trials (n) and the probability of success (p) (Ross, 2010). A common example of a Bernoulli trial encountered in everyday life is a coin flip; there are only two outcomes8—"heads" or "tails." In the case of a fair coin, the probability of obtaining heads is 0.5, as is the probability of obtaining "tails." The outcomes of a large number of coin flips are distributed as a binomial random variable with the probability of "success" equal to 0.5 (Ross. 2010). Unlike this coin flip example, the underlying probability of a skeletal trait being present in the population is unknown; this is precisely what is being estimated using a large sample of independent skeletal "trials" and logistic regression.

To model the probabilistic relationship between the independent variable (age) and the dependent variable (trait score) a mathematical function is needed. Here the logit—the natural log of the odds—is used (Equation 3.0).

$$logit(p) = ln(odds) = ln\left(\frac{p}{1-p}\right) = ln\left(\frac{probability\ of\ trait\ present)}{probability\ of\ trait\ absent}\right)$$
 (3.0)

-

⁸ Excluding the improbable case where the coin lands perfectly on its side.

The logistic regression is equivalent to a linear combination of the independent variables. So, in this case, the natural log of the odds is equal to a constant (β_0) plus a coefficient (β_1) that is multiplied by age (Equation 3.1)(Agresti, 2007a). The model constant (β_0) and the coefficient for age (β_1) are found through maximum likelihood estimation. Here it is sufficient to say that this is accomplished by computer software iteratively testing different parameters until the values that maximize the likelihood function are found. If the natural log of the odds of (p) is considered to be the response variable (y), the form of the equation is identical to that of a standard linear regression $y = \beta_0 + \beta_1 x$.

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \varepsilon = constant + (coefficient*age) + error term$$
 (3.1)

The error term (ϵ) is assumed to be independent of age. Importantly, the value for the coefficient (β_1) in this model is in terms of the log odds per year, not in terms of probability. This distinction is important because although the change in the log odds is the same for each interval of age, the probability of a feature being present changes non-linearly over the lifetime. As described in more detail below, it is precisely this probabilistic change throughout adulthood that allows relatively low-information skeletal features to collectively contribute substantially to adult age estimation.

To calculate the probability of a feature being present at each age from the fitted logit model, an additional step is needed. Taking the inverse of Equation 3.1 (i.e., the inverse logit) will produce Equation 3.2 (Agresti, 2007b), which is the general formula for the probability of a trait being present at a particular age. Using the estimated values for (β_0) and (β_1) to calculate the value of the function at each age generates a curve showing the probability of the trait existing across the lifespan. Because the feature scores are binary, subtracting the calculated probability from one will give the probability that a trait is absent.

$$Pr(y = 1 \mid x) = estimated probability of trait being present (1) at age (x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$
 (3.2)

Figure 3.0 shows examples of logistic curves (solid black lines) for simulated trait data (black circles). The black dashed lines show an approximate 95% confidence interval for each curve. The vertical black dotted line indicates the median, or "age-at-transition." Although the plots were generated from simulated data, all curves except for (a) and (f) are similar to those commonly calculated from actual data. Plot (a) is extreme case that never occurs in adult skeletons; it is included to illustrate the principle of logistic regression. All individuals (black dots) under 45 years of age do not have the trait, while all individuals above this age do. The fitted logistic curve (solid black line) predicts that sometime after turning 45 the trait begins to form, and by age 46 all individuals will have the feature. The 95% confidence interval (black dashed lines) is narrow and essentially vertical because there is complete separation of the data between 45 and 46 years. Plot (b) shows a curve similar to those often seen for traits that occur in the late teens, twenties, and early thirties such as medial clavicle epiphysis fusion. In this case, the slope of the curve is steep and the median of the fitted regression (vertical dotted line) indicates the age at which it is equally likely for the trait to be absent or present. Because these curves provide probabilistic information for all of adulthood, even curves that span half, or more, of the adult lifespan, such as d, e, and f, are potentially useful.

Three of the plots from Figure 3.0 (c, d, and f) are shown again in Figure 3.1 (top row) with plots showing the transition curves for both the absence and presence of a trait (bottom row). In this direct comparison, it is easier to see that the estimated age-at-transition (vertical dotted line, top row) is the age at which a trait is equally likely to be present or absent (horizontal dotted line, bottom row). Because each feature is binary, as the probability of the trait being absent decreases, the probability of the trait being present must increase accordingly. Although individually each curve provides only broad age information, these curves become useful when transitions that collectively occur in all parts of the adult lifespan are identified and assessed in large skeletal samples, as will be described further in Chapter 4.

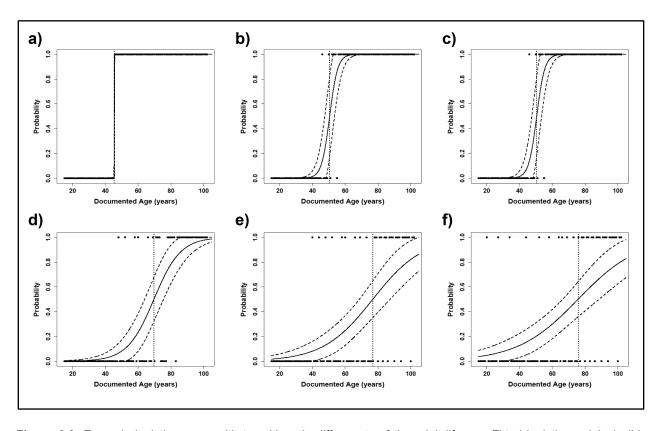


Figure 3.0. Example logistic curves with transitions in differ parts of the adult lifespan Fitted logistic models (solid black lines) with approximate 95% confidence intervals (dashed black lines) and estimated ages-at-transition (vertical dotted lines) based on simulated trait data (black dots).

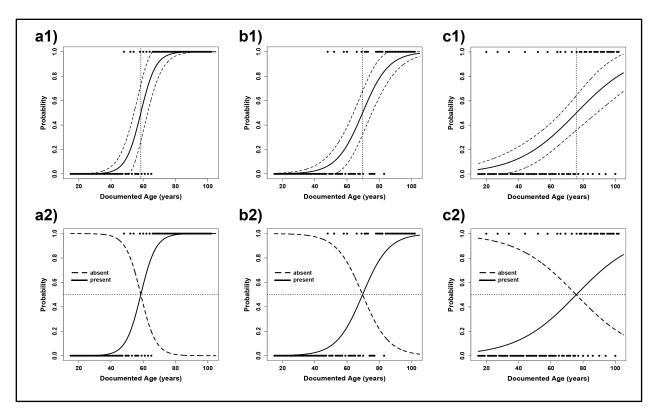


Figure 3.1. Example transition plots from Figure 3.0 (a1, b1, and c1), shown with the logistic curves for both the presence and absence of a trait (a2, b2, and c2). The estimated age-at-transition (top row, vertical dotted line) is the point at which there is an equal probability of a trait being absent (bottom row, dashed curve) and present (bottom row, solid curve).

Data Collection

Well-documented skeletal collections are in high demand from researchers around the world. For most collections, applying for access to skeletal samples must take place well in advance of data collection and scheduling and negotiations can take upwards of several months. Once the terms of use are negotiated and access is granted, the actual process of travel and data collection is a time consuming and costly endeavor.

Between 2012 and 2014, I collected a large data set from four well-documented North American skeletal collections as part of Phase 1 (Table 3.0). However, before primary data collection commenced, seven rounds of preliminary data collection, analysis, and procedural refinement were undertaken at the William M. Bass Donated Collection to identify and refine new skeletal traits and improve scoring procedures before resources were invested in smaller and more geographically distant skeletal samples. The first two rounds of preliminary data collection were completed by George Milner prior to 2012, with the additional five rounds completed as part of this dissertation.

Table 3.0. North American skeletal collections used in Phase 1

Skeletal Collection	Location	Collection Origin	Ancestry	Place of Birth
WM Bass Donated	Knoxville, TN	donated/forensic	white	USA
Maxwell Museum	Albuquerque, NM	donated/forensic	white	USA
UI-Stanford	Iowa City, IA	anatomical donations	white	Europe, USA
JCB Grant	Toronto, ON	anatomical donations	white	Canada

Pilot study & preliminary data collection

In August 2012, a one-week pilot study, the first preliminary round of data collection for this dissertation, was conducted at the Bass Collection. This trip had four main objectives: 1) to become familiar with general age-related changes in the adult skeleton, 2) collect data on 51 features from the original list generated by the research team, 3) generate a list of new potentially age-informative traits, and 4) establish a baseline for a novice researcher's ability to subjectively estimate age from the skeleton. Objectives one through three are key components of trait identification and refinement, while objective four relates to a different component of this research that will be discussed in more depth in Chapter 5.

Before arriving at the Bass Collection, an age-balanced sample of 100 individuals drawn from five-year age categories was selected from all available white males and females in the collection. Although a precisely age-based sample is not critical for this preliminary work, a sample selected in this way ensures that the entire age range represented in the collection will be observed. Prior to the start of data collection any demographic information, including age and sex, present on the outside of the box for each individual in the preselected sample was covered. If an individual could not be scored, either because the remains were unavailable for study or the majority of the features could not be observed, another individual from the same five-year age category was selected and mixed into the remainder of the preselected sample to ensure blind data collection. On days one through four, 105 individuals were scored for 51 unique skeletal features. On days four and five, an additional 19 individuals were selected to represent the entire adult lifespan and were examined in detail to search for additional traits that potentially exhibit age-related variation.

Data were collected for each skeleton on paper forms and entered into Excel after returning from the field. Traits with more than two categories were recoded electronically as dichotomous variables. For example, a trait with three possible scores—absent, greater than or equal to one-half (≥1/2), or complete—would be analyzed as two separate binary features: absent versus (≥1/2) and (<1/2) versus (≥1/2). The binary scores were analyzed using both a generalized linear model (GLM) fitted with the glm() function, a standard part of the program R, and a generalized additive model (GAM) generated with the gam() function in the R package mgcv (R Core Team, 2012; Wood, 2004; Wood, 2006). Graphical visualizations of both models were used to evaluate the potential age-related data present in each trait variant. Although the construction of the two models differs significantly, their graphical interpretations are similar. At this preliminary stage of analysis, it is simply important that the binomial logit model produces a single constant and coefficient that describe how the probability of having a feature changes over the adult lifespan. In contrast, the gam() function fits a curve to different segments of the data and, depending on the degree of smoothness selected, allows for a more complex

function. In some cases, the bumps and dips allowed for in the generalized additive model may represent real and ultimately important underlying features of the data. However, in small samples these model features may just as easily represent noise in the data. Further discussion of these models and their effects on maximum likelihood age-at-death estimates are found in Chapters 4 and 5.

Evaluating Fitting Models Using Transition Plots

Once a model is generated, the results can be classified as one of three patterns: 1) the binary trait scores had a strong relationship with chronological age, 2) scores showed a possible relationship to age with a general trend of older individuals having older character states, or 3) scores had no apparent relationship with age with both trait categories being evenly distributed across the entire lifespan. The potential value of each trait was assessed by looking at both the logistic and general additive model curves while taking into account the age distribution of individuals with each score. Based on these data, features were 1) accepted with no modifications, 2) modified to better capture biologically meaningful categories, or 3) eliminated from future study. Because samples sizes in all of the preliminary investigations were small—at most several hundred individuals—decisions to revise or eliminate features relied more heavily on the distribution of trait scores than on the fitted curves. Trait categories were eliminated if they contained no age-related variation or could not be reliably scored, even after several rounds of refinement, and additional categories were added as they were identified. Trait definitions were refined and elaborated based on personal experience with the features to improve scoring reliability. In general, traits were first tested using relatively nebulous binary categories, which were then divided into multiple, more explicitly defined variants as the extent of trait variation became clearer with experience. Ultimately, many of the variants were collapsed to reform binary trait categories with well-defined thresholds.

Figure 3.2 shows a schematic overview of the trait evaluation, refinement, and selection process. The cycle begins with trait identification and definition (Figure 3.2, upper left). In each iteration of data collection, traits with strong age-related patterns are retained, while traits that provide little to no age information are revised for one or more rounds of data collection before being eliminated if strong age-related patterns fail to emerge. The process ends when a suite of traits that provides information across the entire adult lifespan has been identified (Figure 3.2, upper right). Table 3.1 shows the samples used, the number of traits and transitions identified, and the features eliminated after each round of analysis.

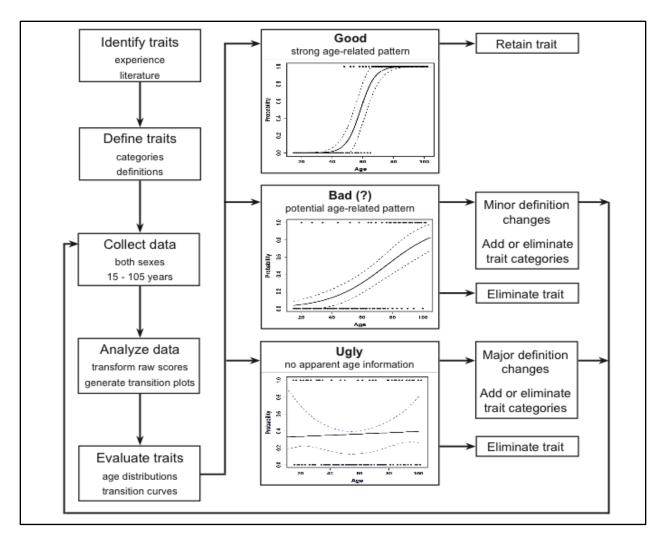


Figure 3.2. Schematic of the trait identification and refinement process in Phase 1, beginning with trait identification and definition (upper left) and ending when a suite of traits that provides information across the entire adult lifespan has been identified (upper right).

Table 3.1. Summary of the Phase 1 trait identification & refinement process

		Skeletal Sample Traits						
Collection	Date	N	M	F	New	Used	Transitions	Eliminated
Bass	08/12 - 08/17/12	124	66	58		50	79	7
Bass	05/12 - 05/24/13	122	60	62	26	69	176	9
Bass	06/17 - 06/21/13	101	44	57	2	62	137	3
Bass	07/22 - 08/02/13	196	91	105	7	66	162	6
Bass	08/05 - 08/30/13	500	272	228	0	60	151	8
Maxwell	04/17 - 05/02/14	170	101	69	1	53	104	0
UI-Stanford	07/14 - 07/18/14	149	129	20	0	53	105	0
Grant	08/05 - 08/15/14	191	175	16	0	53	105	8
Final		1553*	938	615	45		80	

^{*} representing 1,119 unique individuals from four reference collections

In the first round of preliminary data collection (i.e., the pilot study), twenty-six new skeletal features were identified and defined for future data collection and seven traits were eliminated (Table 3.1). The first four rounds of data collection and analysis followed the same process of data collection, analysis, and trait refinement described above. The fifth round of data collection lasted six weeks and included both the final round of preliminary trait refinement and the beginning of primary data collection. Data were collected for two weeks and analyzed in week three when a final set of features to be assessed in all four of the samples was selected. These features were evaluated at the Bass Collection in weeks four through six, and at the three remaining collections during the following year. Although 1,553 sets of data were examined in Phase 1, these data represent only 1,119 unique individuals; some skeletons from the Bass Collection were included in more than one preliminary sample because there are a limited number of individuals in the youngest and oldest age categories.

Primary data collection

The five rounds of preliminary data collection focused on refining features so they would be quick and easy to score. Whenever possible, the number of variants was reduced or the thresholds needed for a trait to be present were adjusted to clarify scoring descriptions with the intention of reducing observer error (Shirley & Ramirez Montes, 2015). Consideration was also given to selecting features that are often preserved in archaeological and forensic settings, which increases the probability that age estimates can be generated from partial remains.

All individuals in the primary sample were born in the nineteenth or twentieth centuries, have well-documented age-at-death, and are of European (i.e., white) ancestry. Because of their common ancestry, the individuals likely share some measure of underlying genetic similarity, but were intentionally chosen to represent as heterogeneous a group as possible in other respects. To address the potential issue of secular change in skeletal aging, the primary reference sample is divided into two "death cohorts": 1) pre-1953—individuals who died between 1924 and 1952 and 2) post-1981—individuals who died between 1982 and 2013. Although a somewhat messy distinction, the time period just following WWII is generally regarded as the beginning of a new age of change for epidemiology, medical technology, public sanitation, and food processing and distribution (Susser, 1985). The division into two death cohorts separates the sample into groups that, on the whole, lived with substantively different nutrition and ability to combat infectious disease.

Individuals in the pre-1953 death cohort lived most or all of their lives during a time when infectious diseases such as tuberculosis, pneumonia, influenza, typhoid, dysentery, diphtheria, pertussis, measles, polio, and syphilis were not uncommon (Armstrong, Conn, & Pinner, 1999). Although a number of vaccines were developed prior to 1900, including those for smallpox, cholera, rabies, and plague, they were not used widely enough to control disease for at least several decades into the twentieth century (Centers for Disease Control and Prevention, 1999). Deaths in the pre-1953 cohort were dominated by infections, accidents, and other ailments, such as strokes, heart disease, cancers, liver disease, and diabetes.

In the years following WWII, medical technology improved and infectious diseases declined with the discovery and wide distribution of new antibiotics, the development of vaccinations and associated distribution programs, and better sanitation practices. Individuals in the post-1981 death cohort had the benefit of sulfonamides, antibiotics, and antimycobacterials that were first used to treat many diseases in the 1930s and 1940s (Armstrong et al., 1999). Their population-level disease exposure was also greatly reduced by the development and widespread

application of vaccines for pertussis (whooping cough) (1926), tetanus (1927), polio (1955), measles (1964), mumps (1967), and rubella (1970) (Centers for Disease Control and Prevention, 1999; Cutler, Deaton, & Lleras-Muney, 2006). Other medical advancements, including safer forms of anesthesia, fluid and electrolyte replacement therapy, blood typing and blood banks, chemotherapy, and cardiac surgery with the use of a heart-lung bypass machine, also became standard parts of the medical arsenal in the second half of the twentieth century (Fou, 1997; Guyer, Freedman, Strobino, & Sondik, 2000; Susser, 1985). After 1950, the United States also saw the widespread adoption of water treatment practices, including filtration and chlorination, the formation of local and state level health departments with programs for solid waste disposal and public hygiene education, and new social programs aimed at improving the lives of the poor and elderly (Binstock, 1991; Cutler & Miller, 2005; Guyer et al., 2000; Warner, 2012). Major changes to the way food was processed and transported also occurred. These included improvements in heat processing and canning, the ability to easily freeze and refrigerate food, and an explosion in the ingredients and additives available to food manufacturers. Combined with the substantial development of the US Interstate Highway system by the mid-to-late 1960's, these developments brought substantive changes to the average American diet (Weber, 2012; Welch & Mitchell, 2000).

Although exact life-history information for each individual is unknown, it is safe to assume that the death cohorts differ simply by virtue of variation in available foods, advancements in medical care, and changing social and economic conditions in the different locations over the one-hundred and fifty years when these individuals lived and died. Furthermore, it can be reasonably assumed that, collectively, these samples encompass a large fraction of the variation that can be expected in any European population. This approach of combining multiple samples is the most conservative one because it creates as large and diverse a reference sample as possible within a single broad ancestral category (Konigsberg, Herrmann, Wescott, & Kimmerle, 2008; Milner & Boldsen, 2012b; Townsend & Hammel, 1990). In many forensic situations, and essentially all archaeological ones, it is impossible to identify which regional or population-specific reference sample would be the most appropriate, so a widely applicable combined reference sample is desirable. While this approach is likely to decrease the precision of individual age estimates, the increase in applicability without sacrificing accuracy is desirable.

Skeletal collections

The William M. Bass Donated Skeletal Collection at the University of Tennessee was established in 1981 and new individuals are added as they become available. The Bass Collection is currently composed of over 1200 individuals, mostly of European ancestry, with the vast majority born after 1940; it is the largest, well-documented skeletal collection of modern people in North America. All individuals have been donated to the facility by individuals prior to their deaths, by family members, or by the local Medical Examiner (Forensic Anthropology Center, 2012). Age, sex, ancestry, and cause of death are available for almost all individuals, and additional data, such as birth information, medical history, occupation, socioeconomic status, and habitual activities are available for most individuals after 1999.

_

⁹ Although it was not officially eradicated in the US until 1977, smallpox was not a significant public health concern in the US in the late nineteenth and early twentieth centuries; however, outbreaks did occasionally occur, particularly among immigrant populations and the poor (Centers for Disease Control and Prevention, 1999; Colgrove, 2006). The smallpox vaccine was not produced in Ontario, Canada, until 1886 and outbreaks were common in this area until they were brought under control in the 1920s (Archives of Ontario, 2012a, 2012b). Although the extent to which the individuals in the first death cohort were impacted by smallpox is unknown, the samples from the UI-Stanford and Grant collections represent populations that were potentially at risk for exposure and infection.

The Maxwell Museum Donated Skeletal Collection at the University of New Mexico was established in 1984 and is an active collection with new individuals added as donations are made. The collection currently contains approximately 300 individuals who died in the Albuquerque area and were donated by the individual before death, by the deceased's family, or by the Office of the Medical Investigator (Maxwell Museum of Anthropology Laboratory of Human Osteology, 2010). In addition to basic demographic data, including age, sex, self-identified ancestry, and cause-of-death, health and occupational data have been requested for donations made since 1995.

The University of Iowa-Stanford Collection is composed of over 1,100 individuals who died in the San Francisco Bay-area and were donated for medical study in the 1920s through the early 1950s. The collection was officially transferred from Stanford University Medical School to the University of Iowa and the Office of the State Archaeologist (OSA) in 1998 (University of Iowa & Office of the State Archaeologist, 2012b). Between 2000 and 2002, a National Science Foundation Grant (NSF) for the conservation of the collection funded the reassessment of donation documentation and the rehousing of individuals into archival-quality polypropylene storage containers for long-term curation (University of Iowa & Office of the State Archaeologist, 2012a). The information available for many individuals is extensive, particularly considering that it was collected in the first half of the twentieth century, and includes birth date, birthplace, death date, age, occupation, residence at death, birthplaces of parents, and cause of death. This collection is a rare example of well-documented sample of individuals who lived and died at a time when antibiotics and most modern medical treatments were not yet available. This makes the sample particularly attractive for the investigation of features that will be useful for the estimation of age in individuals from the historic and prehistoric past.

The J.C.B. Grant Collection is composed of 202 individuals who were donated to the University of Toronto Department of Anatomy between 1928 and the early 1950s. The majority of individuals were donated from local hospitals and welfare institutions, and many were transients, migrant workers, or recent immigrants (Bedford et al., 1993). The number of individuals donated was originally much larger but, in 1948, all individuals whose age-at-death could not be verified through checks with vital statistics records, hospital records, or personal history provided by the individual prior to death were removed from the collection (Bedford et al., 1993). Documentation for each individual includes name, sex, age-at-death, and cause of death. The individuals in this collection lived contemporaneously with those in the UI-Stanford collection, but are a separate North American sample with potential differences in genetic composition, habitual activities, socio-economic status, and diet.

All individuals who met the selection criteria (i.e., well-documented ancestry, sex, and age-at-death) were analyzed from the Grant and UI-Stanford collections. For the Bass and Maxwell Museum collections, an attempt was made to select an equal number of individuals from all five-year adult age categories. However, essentially all skeletal collections have fewer individuals in the youngest and oldest age categories than in middle age, so individuals between 50 and 80 years of age are over-represented in the final skeletal sample. Females are also less frequent in many documented skeletal collections, particularly those formed at medical schools in the early-to-mid twentieth century, so the final sample is more heavily weighted towards males. Table 3.2 shows the collections, sample sizes, and range of birth and death years in each of the samples. Figure 3.3 shows the age and sex composition of the combined sample with fitted normal distributions. Both distributions are slightly left skewed with females being slightly older on average (N=333, mean=65.7 years, SD=17.7) than males

(N=677, mean=61.7 years, SD=16.8). Figure 3.4 shows the age and sex distribution of the sample analyzed from each of the four collections and in the two death cohorts.

	Table 3.2. Phase 1	primary	reference	sample	death	cohort	informatio
--	--------------------	---------	-----------	--------	-------	--------	------------

Death Cohort	Collections	Birth Years	Death Years	N	М	F
1: (Pre-1953)	JCB Grant	1836 - 1922*	1928 - 1949	191	175	16
1. (1 16-1933)	UI-Stanford	1842 - 1907	1924 - 1952	149	129	20
2: (Post-1981)	WM Bass	1893 - 1990	1982 - 2012	500	272	228
2. (1 031-1301)	Maxwell Museum	1887 - 1975	1984 - 2013	170	101	69
Combined		1836 - 1990	1924 - 2013	1010	677	333

^{*} birth years for the Grant Collection were estimated by subtracting the documented age of each individual from the range of years the collection was in operation

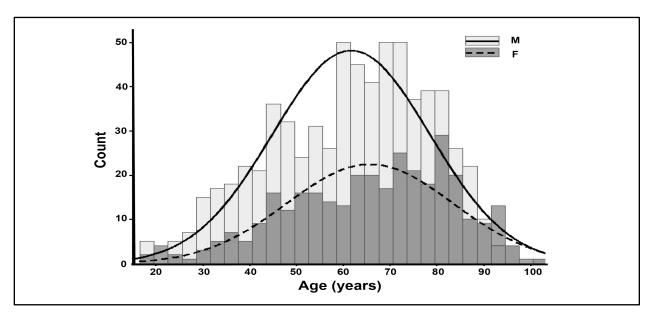


Figure 3.3. Age distributions of males and females in the Phase 1 primary reference sample.

Although usually not viewed as an advantage, there is reason to believe that because the statistical approach used here does not depend directly on sample composition, having more individuals 50 and 80 years of age can only help to more effectively characterize the variability seen in this part of adulthood. This is because, on average, the more variability there is in a population, the larger sample you will need to effectively characterize it. It is commonly believed that skeletal variation increases throughout life making it impossible to estimate age at the upper reaches of adulthood (Franklin, 2010). However, published work showing the ability of existing TA and an experienced observer to predict age throughout adulthood indicates that this is likely untrue (Milner & Boldsen, 2012c). Additional data collected by the research team, including portions of this dissertation, support the notion that skeletal variation in age indicators is relatively low in the youngest ages where it is more tightly controlled by developmental processes, and also lower in the very old, potentially as a result of selective mortality. Although our discussion of the effects of selective mortality on skeletal variation is speculative at this

point, a growing body of data from this dissertation and the NIJ project indicates that this process could be significant for our understanding of adult aging. These effects could potentially play an important part in increasing variation in middle age where existing TA and expert age assessments perform the worst. Further discussion of potential selective mortality signals in the data and their impact on age estimation can be found in Chapter 4.

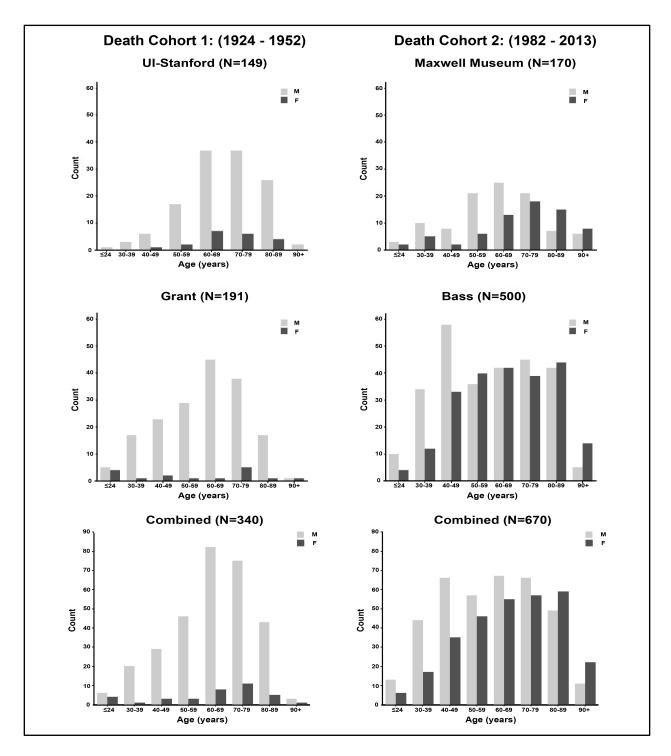


Figure 3.4. Age distribution, by sex, of the individuals evaluated from each skeletal collection and death cohort.

Trait Selection

In Phase 1, 100 traits, with almost 250 variants were investigated in eight rounds of data collection. A primary reference sample of 1,010 individuals was used to assess 53 traits using both logistic and generalized additive models. All data were analyzed collectively and by collection, sex, and death cohort. Figure 3.5 shows the key features of the transition plots used to evaluate the age-informative potential of each feature, including the age distribution of the trait scores, the shape of the regression curves, and the estimated age-at-transition.

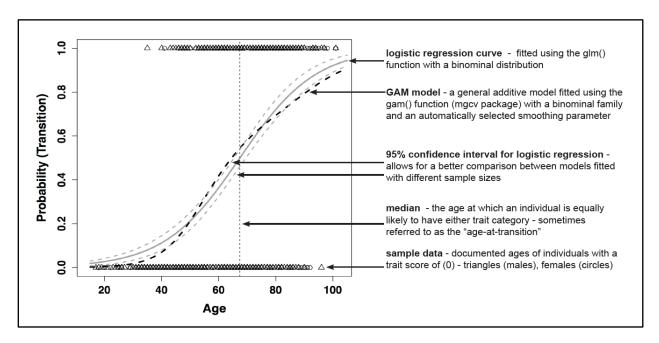


Figure 3.5. Key features of the transition plots used to evaluate Phase 1 traits.

With the exception of two rib traits where sample sizes were extremely small, five plots were generated for each transition between binary trait categories: 1) total combined sample, 2-3) combined sample separated by sex, and 4-5) sample separated by death cohort. The combined sample includes all individuals from the four samples scored for the trait of interest. Separate models were then generated for males and females. The remaining two plots compare the trait scores of individuals in death cohort 1 (pre-1953) and death cohort 2 (post-1981). Because of the small number of females who died in the first death cohort, males and females are not analyzed separately. Table 3.3 shows the sample size used to generate these models for each trait.

¹⁰ The exact number of traits and variants investigated is difficult to determine because of the overlapping nature of some features and the iterative nature of the trait refinement process (i.e., how much change must take place in a trait definition for it to be considered a "new" variant). Many areas of the skeleton were investigated for multiple features, including shape and textural changes, ossification of soft tissue structures, and combinations of these changes, often with several different metric and non-metric thresholds evaluated.

Table 3.3. Phase 1 primary reference sample trait analysis sample sizes

	Combined Sample			Death Cohort 1 (pre-1953)			Death Cohort 2 (post-1981)		
Trait	F	M	N	F	M	N	F	M	N
hyperostosis frontalis interna (HFI)	76	229	305	17	135	152	59	94	153
parietal depression	304	484	788	17	135	152	287	349	636
occipital condyle lipping	286	449	735	8	109	117	278	340	618
R1 costal face	240	366	606	8	63	71	232	303	53
R2 rim edge profile	190	232	422	1	3	4	189	229	41
R310 rim edge profile	210	338	548	5	58	63	205	280	48
R310 body thickness	301	474	775	10	110	120	291	364	65
sternal central dorsal ridges	142	221	363	10	76	86	132	145	27
trapezium lipping	255	570	825	34	280	314	221	290	51
cervical lipping	300	493	793	13	134	147	287	359	64
thoracic lipping	301	506	807	13	141	154	288	365	65
lumbar lipping	289	498	787	12	138	150	277	360	63
cervical candlewax	301	493	794	12	134	146	289	359	64
thoracic candlewax	299	502	801	13	138	151	286	364	65
lumbar candlewax	292	501	793	12	136	148	280	365	64
C1 lipping	296	482	778	12	139	151	284	343	62
C1 eburnation	293	477	770	11	135	146	282	342	62
L5 superior margin	267	477	744	9	129	138	258	348	60
L5 inferior margin	260	467	727	8	120	128	252	347	59
S1 superior margin	258	470	728	13	143	156	245	327	57
S1-2 fusion	296	523	819	17	173	190	279	350	62
sacroiliac joint fusion	304	562	866	19	208	227	285	354	63
clavicle medial epiphysis fusion	284	511	795	23	196	219	261	315	57
clavicle medial bone growth	276	433	709	21	162	183	255	271	52
clavicle lateral macroporosity	280	515	795	23	188	211	257	327	58
scapula glenoid fossa margins	307	577	884	25	218	243	282	359	64
humerus medial epicondyle	304	606	910	25	250	275	279	356	63
humerus lateral epicondyle	312	606	918	27	249	276	285	357	64
humerus lesser tubercle lipping	281	560	841	25	230	255	256	330	58
humerus lesser tubercle bumps	281	560	841	25	230	255	256	330	58
radius medial crest	301	631	932	32	281	313	269	350	61
femoral fovea margin lipping	266	571	837	28	260	288	238	311	54
femoral head surface	278	594	872	26	262	288	252	332	58
trochanteric fossa exostoses	277	505	782	18	172	190	259	333	59
medial trochanteric fossa exostoses	264	564	828	22	237	259	242	327	56
acetabulum posterior margin lipping	300	599	899	26	251	277	274	348	62
acetabulum articular surface bone growth	303	602	905	27	253	280	276	349	62
pubic symphyseal collar	288	574	862	20	231	251	268	343	61
ischium superior margin spur	313	608	921	27	255	282	286	353	63
ischium bone growth	312	604	916	25	249	274	287	355	64
AIIS exostoses	289	591	880	23	245	268	266	346	61
humerus weight	327	662	989	35	293	328	292	369	66
tibia weight	322	649	971	33	287	320	289	362	65
innominate weight	313	484	934	28	271	299	285	350	63
calcaneus weight	313	621	784	18	121	139	282	363	64

Appendix B contains the transition plots for each of the 45 features selected, including fitted models for the combined sample and subsamples separated by sex and time period of death. The left side is shown for bilaterally scored features unless otherwise noted. The vast majority of features have only two variants (absent and present), so only one group of transition curves are shown; however, several traits have three or four variants and thus have multiple sets of curves. For example, the medial clavicle has four variants (not fused, partially fused, remnant line, and completely fused), so the results for three binary transitions are shown. Because correlated traits are problematic, only one transition per feature is used in the final age-estimation procedure tested in Phase 3.

Trait examples and interpretation

Traits that can be universally applied are preferred to those that require sex- or populationspecific standards, particularly for archaeological applications where an appropriate reference sample is often unavailable or unknown. The presence of raised, irregular bone growth on the lateral margin of the lesser tubercle of the humerus is an example of a trait with almost ideal characteristics (Figure 3.6). The humerus is typically well-represented in skeletal collections and the margin of the lesser tubercle is often preserved, even in fragmented remains. The combined sample size is sufficiently large and adequately covers the entire adult lifespan, resulting in a well-behaved (smooth) probability curve with narrow confidence intervals throughout adulthood. The trait also makes a full "transition" over the course of the lifespan. In other words, essentially all individuals under 40 years of age do not have the feature, while all those over 90 do. This means that both the absent and present categories can provide useful age information over the entire lifespan. Additionally, the curves for males and females have almost identical median ages and the two death cohorts are similar. The wider confidence interval and slightly broader slope for death cohort 1 is likely the result of too few individuals under the age of 50 in the reference sample. This issue influences of all curves generated for death cohort 1, so the age distribution of trait scores becomes particularly important for comparing traits between the two death cohorts.

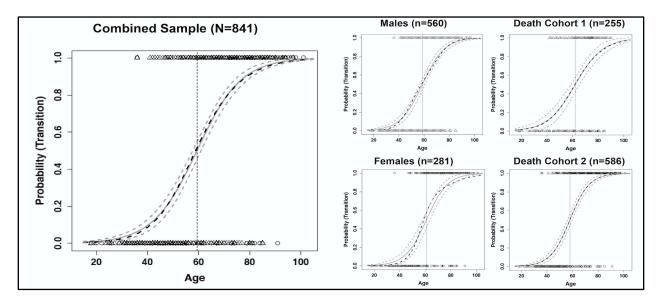


Figure 3.6. Transition plots for irregular bony growth on the margin of the lesser tubercle of the humerus.

The fitted models for anterior superior iliac spine exostoses, shown in Figure 3.7, are similar to those in Figure 3.6 except that slightly larger differences can be seen in the sex- and death cohort-specific models. The sex-specific sample sizes are reasonably large, 591 and 289 for males and females respectively, with adequate coverage of the entire lifespan. Thus, the differences seen in this case are unlikely to be the result of sampling. The final age-estimation method developed by the NIJ-research team for dissemination to the osteological community will have the option of using a combined or sex-specific reference sample. A combined sample would contain the variation seen in both males and females and therefore likely produce slightly broader (i.e., less precise) age estimates. The decrease in precision, however, may be justified by an increase in the applicability of the method for fragmented remains or isolated skeletal elements where the estimation of sex is not possible with a high degree of confidence. The issue of sex-specific models is further investigated in Chapter 5.

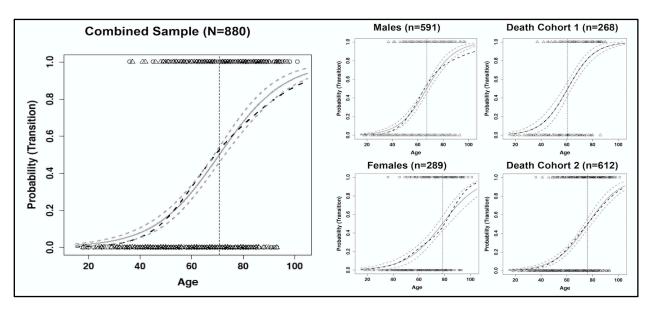


Figure 3.7. Transition plots for the presence of exostoses on the anterior inferior iliac spine (AIIS).

The final trait example is one with a different, but relatively common, pattern. The presence of candlewax on the vertebrae—shiny, sclerotic bone with the appearance of melted wax, often called diffuse idiopathic skeletal hyperostosis (DISH) when present on four or more consecutive vertebrae—is an example of a feature that occurs relatively infrequently, but that may contain some age information (Figure 3.8).

In cases with low overall sample size, such as cervical candlewax, it is especially important to look at the overall age distribution of the trait scores in each sample before interpreting the fitted models. In the combined sample, it appears as though the presence of candlewax on at least one cervical vertebra provides some information after around age 60. However, unlike the humerus lesser tubercle margin where both categories provide age information, the absence of the trait provides little data about age; individuals without the feature range from 15 to 101 years of age—the absolute limits of the age distribution. Looking at males and females separately, there is a slight difference in the age distributions. Candlewax appears at an earlier age in males. Interestingly, individuals with the feature also "disappear" from the sample at an earlier age. Because of the relatively small sample sizes involved at the upper ends of the lifespan, it is possible that this difference can be explained by random chance in sampling.

However, the possible influence of selective mortality is also being explored by the research team with a larger and more diverse sample as part of the larger NIJ-funded project. In the case of cervical candlewax, it is possible that males who develop the feature may have an increased risk of death relative to their age-matched counterparts who do not. Although speculative at this point, given this trait's associations with adult-onset diabetes, obesity, hypertension, vascular and purine metabolism disorders, and potential to cause severe spinal cord injury from relatively minor trauma (Hannallah, White, Goldberg, & Albert, 2007; Kiss, Szilagyi, Paksy, & Poor, 2002; Poelstra, 2013), an increased risk of death, particularly for males, would not be surprising. There is also difference between the two death cohorts. In this case, however, the two individuals with candlewax in death cohort 1 fit comfortably within the age distribution for the feature in the combined sample. Therefore, it is likely that the differences are likely the result of sample size and not secular change (i.e., change over time). While it is possible that cervical candlewax has the same age distribution in both death cohorts, a much larger sample would be needed to investigation this possibility.

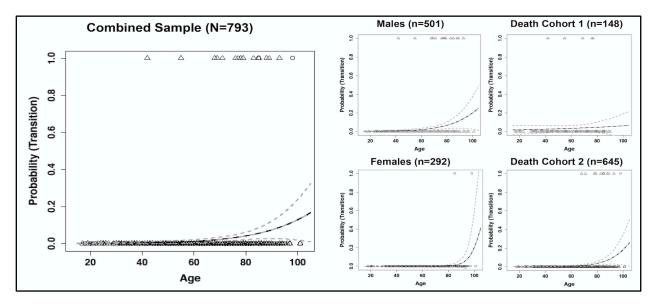


Figure 3.8. Transition plots for the presence of candlewax on at least one cervical vertebra.

Final Phase 1 Traits

Figure 3.9 shows the distribution of all features investigated in Phase 1 and the 45 chosen for further investigation. In Phase 2 of this project, this list of features is combined with additional traits identified by members of the research team and tested in three modern samples from South Africa, Thailand, and Portugal to assess their utility in other populations. Features that show similar patterns among the temporally and geographically diverse samples in Phases 1 and 2 are incorporated into the new age-estimation procedure that will be tested in Phase 3.

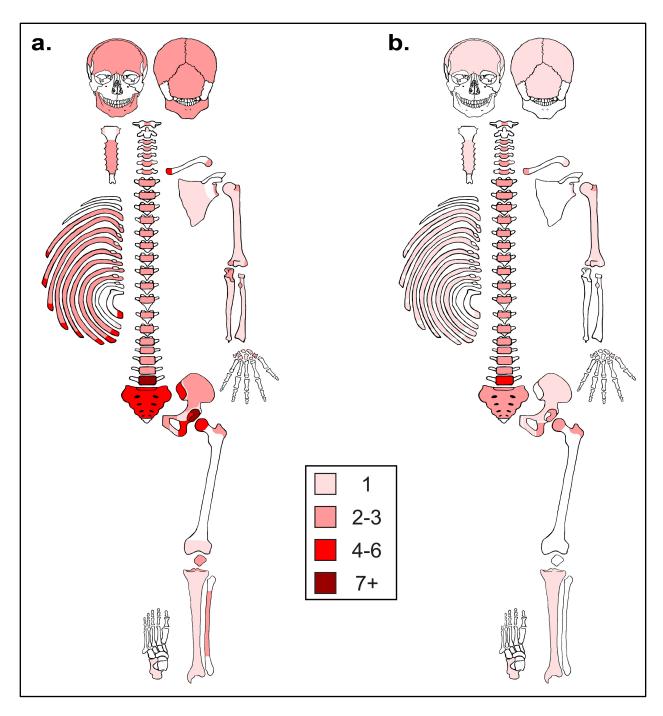


Figure 3.9. (a) Distribution of the approximately 100 features investigated in Phase 1 and (b) the location of the final 45 features identified as potentially informative for adult age estimates.

CHAPTER 4: PHASE 2—COLLABORATIVE TRAIT INVESTIGATION & REVISED TRANSITION ANALYSIS METHOD DEVELOPMENT

Accurate and precise estimates of age-at-death are often critical when individuals are represented only by skeletal remains. Unfortunately, practically useful age estimates for adults—those where an individual can be assigned confidently to a relatively narrow interval with a high degree of confidence—are beyond what standard methods can provide (Milner & Boldsen, 2012b). Existing methods most often yield biased point estimates of age that are accompanied by ranges that span several decades to virtually all of adulthood. The age intervals assigned to skeletons often cannot narrow a search to certain categories of missing people or provide support for individual identifications and increasingly fail to meet rigorous court-imposed standards (Garvin & Passalacqua, 2012; National Research Council, 2009). Although many have tried to address these issues¹¹, accurate age estimates for adult skeletons remain elusive. Furthermore, little can be said with confidence about the age of people who live beyond about 50 years, a group who currently represents just over a third of the US population (Howden & Meyer, 2011).

Beginning in 1996, Drs. George Milner and Jesper Boldsen began collaborative work on the skeletal and statistical components of the adult skeletal age-estimation problem. The primary objective of their work was to address the systematic bias in age estimates from adult skeletal remains. Although small portions of this research were funded in conjunction with other projects, anthropological funding agencies typically do not support method development. The single research proposal focusing solely on existing Transition Analysis (TA) and experienced-based age assessments, submitted by Milner, was unsuccessful. The primary objection provided when the grant was rejected was that existing age-estimation methods work reasonably well in the younger half the adult lifespan, and few, if any, people over about 50 years of age exist in archaeological samples. Therefore, improved age estimates would not make much of a difference in our interpretations of the past. So, for nearly two decades, this work remained unfunded and collaborators and students worked on aspects of the project essentially asynchronously as time and money permitted.

In December 2014, after several submissions and the addition of two pilot studies¹², a National Science Foundation Biological Anthropology Doctoral Dissertation Research Improvement Grant was awarded to Getz (DDRIG # 1455810; G. Milner, Advisor). At that time, the NSF grant was thought to be the primary means by which the research team could operationalize new aspects of their work. Getz would work closely with Milner to identify and refine skeletal features and evaluate a large sample of known-age skeletons from North America. The research team would collectively continue work on method development using Getz's reference data, and the newly developed technique would then be tested on additional known-age European samples and applied to archaeological samples by Getz as a proof of concept for the procedure's application to paleodemographic analyses. Boldsen and Ousley (Mercyhurst University),

¹¹ For numerous examples of published methods, validation studies, and population-specific applications based on the cranial sutures, pubic symphysis, sacroiliac joint and ribs, see Appendix A.

¹² The first, by Getz, provided a back-to-back comparison of commonly used age-estimation methods, existing TA, and experience-based estimates using modern known-age skeletons from the Maxwell Museum Donated Collection (N=55). Experience-based estimates were used as a proxy for what is expected by incorporating a large suite of skeletal features into the existing TA framework. The second pilot study, presented maximum likelihood estimates of age-at-death generated by Boldsen using preliminary transition analysis for 15 traits using data collected by Milner at the Bass Collection (N=256).

primarily responsible for the statistical and software development aspects of the larger project respectively, offered their expertise and continued support at no cost to the NSF grant.

In April 2014, research team founders Milner and Boldsen, along with Oulsey, replied to an NIJ solicitation specifically for funding basic research to develop accurate, reliable, cost-effective, and rapid methods for the analysis and interpretation of physical evidence for criminal justice purposes (National Institute of Justice, 2014). Shortly after receiving word that the NSF grant had been awarded, the research team was also notified that the NIJ grant to support the much-larger project had also been funded. While the previous anthropological grant had been unsuccessful on the grounds that too few ancient people lived beyond 50 years of age to justify any investment, more accurate and reliable methods were recognized as precisely what are needed for forensic purposes. Milner (PI), Boldsen (Co-PI), and Ousley (Co-PI) were awarded just over a half-million dollars to develop and test a new method of adult skeletal age estimation, as well as to develop associated computer software and disseminate the method through meeting presentations, workshops, and Fordisc (Ousley & Jantz, 2015), the state-of-the-art software used in medicolegal investigations of skeletal remains.

Since its conception, the international research team has solidified and expanded to include additional individuals as its scope developed interdisciplinary demographic, archaeological, and forensic foci. With the funding of the NIJ project, the team now includes researchers from three institutions: The Pennsylvania State University (George Milner & Sara Getz); The University of Southern Denmark (Jesper Boldsen, Svenja Weise & Peter Tarp); and Mercyhurst University (Stephen Ousley). Additionally, Jutta Gampe, a mathematician from the The Max Planck Institute for Demographic Research (Rostock, Germany), is not a member of the NIJfunded research team, but she contributes to projects that are part of the team's broader research agenda. In addition to funds for supplies and travel, the NIJ grant provided support for several members of the research team working on the project. Svenja Weise, previously a Boldsen doctoral student, received a one-year postdoc, while I received two years of support as a research assistant and will continue as a postdoc for one year after completion of this dissertation. As a research assistant. I work collaboratively with Milner to plan data collection trips, select skeletal samples, and create data collection forms and scoring procedures. In addition to collecting skeletal data as a team member, my responsibilities in the field include managing collection paperwork, specimen photography, and documenting changes and discussions related to the trait manual. After returning from the field, my primary responsibilities are to convert the data into electronic format, conduct preliminary analyses, and update the trait manual definitions and images. In conjunction with Milner, I also contribute substantially to the co-authored abstracts, posters, and presentations based on this work. As a post-doc I will continue this collaborative work preparing publications, readying the trait manual and software for dissemination, and conducting at least one training workshop at a national meeting.

This dissertation is only possible because it builds on the two decades of focused adult ageestimation research already conducted by members of the research team, much of which remains unpublished. This chapter provides an overview of the NIJ project, focusing on the aspects of the work that most closely relate to this dissertation, and summarizes work that has been presented at several national conferences by the research team. Because data collection, analysis, and presentation of results are collaborative endeavors among members of the research team, credit is provided where appropriate throughout this chapter.

NIJ Project Objectives

The NIJ project has five primary goals: 1) identify and define new age-informative skeletal features; 2) assess inter- and intra-observer error; 3) establish trait age distributions in populations of diverse ancestry; 4) investigate theoretical and statistical improvements to the existing TA framework; and 5) develop materials for the dissemination of the new TA method to the osteological community, including an illustrated scoring manual and a user-friendly computer program. Compared to prior age-estimation research, this project is unique in both its size and scope.

Although the NIJ project is scheduled to continue until December 2017, all reference data have been collected at this time. Nearly 1,700 skeletons (N=1,698) from four continents have been evaluated by our international team of six osteologists with differing levels of familiarity with age-estimation techniques. Because of the many years of foundational research and collective decades of experience on the research team, this project is able to address three critical areas of methods development – the bones, the math, and the computer software. The project will define and evaluate age-informative traits, refine a mathematically sophisticated procedure to effectively analyze this information, and produce a user-friendly computer program to allow others to easily use the technique. Together, these aim address both the basic and applied research goals of the NIJ grant solicitation for research and development in the forensic sciences. The final result of the project will be a 1) a new method for adult skeletal age estimation based on visual observation of features, and 2) an easy-to-use computer program that can be immediately put into use by forensic practitioners.

The NIJ Project and Dissertation Phase 2

This dissertation and the larger research-team project have complementary objectives. Both projects are focused on the identification, refinement, and documentation of age-related traits throughout the skeleton and evaluation of the accuracy and precision of estimates based on them. While the NIJ project focuses on modern (twentieth century) individuals of differing ancestry from four continents, this dissertation evaluates the range of variation found in one broadly-defined ancestry group over time (seventeenth through twentieth centuries).

The NIJ project contributes to this dissertation as part of Phase 2. Simplied versions of the statistical procedures under development as part of the NIJ project are used to analyze the data collected in Phase 1. These data are then used as a reference sample to test the new TA procedure in Phase 3. At the completion of the NIJ project, data collected by the team, as well as some of the data from Phases 1 and 3 of this dissertation, will be incorportated into a new version of the TA program (Boldsen et al. 2002). This program combines age information from a large number of features with information about population structure to produce age estimates for the entirety of adulthood.

Establishing a baseline for method improvement

The ultimate goal of the NIJ project is to produce accurate, precise, and reliable age estimates for entire adult lifespan. To assess progress towards this goal, a baseline performance level for existing procedures was established during the first round of data collection at the Bass Collection. To avoid the issue of observer error, data were collected by one member of the research team (Milner) for 234 individuals using standard methods for the pubic symphysis and auricular surface, as well as using existing TA. Figure 4.0 shows the results of these techniques. The documented age of each individual in the sample is plotted

against the estimated point age, which is a mean, midpoint, maximum likelihood estimate, or the beginning of a terminal open-ended category, depending on the method. Males are shown in blue and females in red.¹³ The identity line (i.e., the line along which known and estimated age would have perfect correspondence) is shown in each plot.

For all methods except for TA, the estimated age of individuals fall along horizontal lines because age can only be predicted using a relatively small number of fixed categories. TA (Figure 4.0, lower right) is the only method that provides individualized point estimates based on the suite of features present (Boldsen et al., 2002). Not only do the point estimates extend to the extremes of the adult lifespan (15–105 years), but the lengths of the age intervals differ based on the skeletal features available for analysis and the agreement among the age information provided by those features.

It is desirable for both forensic and archaeological applications to have a method where individuals can be assigned to relatively narrow age ranges with a high degree of confidence. Existing techniques, however, typically sacrifice either accuracy or precision at the expense of the other (Buckberry, 2015). The precision (age interval length) and accuracy (known age within the estimated range) of the methods tested at the Bass Collection are summarized in Figure 4.1. All standard techniques either produce extremely narrow ranges that most often do not include the true age of the individual (high precision with low accuracy) or extremely wide ranges that encompass all of adulthood and are often correct (low precision with high accuracy). Neither extreme is useful in forensic or archaeological contexts.

Although TA produces the best compromise between accuracy and precision of the existing methods (Figure 4.1), the results are still insufficient for forensic and archaeological applications. The systematic bias resulting from the failure of standard features to keep pace with chronological age after middle age cannot, in insolation, be overcome by any advanced statistical technique (Milner & Boldsen, 2012b, 2012c). To move forward, statistical improvements must be paired with a broader array of age-informative traits, especially those providing information in middle to old age. The NIJ project aims to generate age estimates with the highest precision and accuracy possible. Milner's experience-based estimates are also shown as a proxy for what the NIJ research team hopes to accomplish when many skeletal traits are incorporated into the new TA procedure (Figure 4.1).

¹³ The Katz and Suchey (1986) revision of Todd shows only males because revised female ranges were not provided.

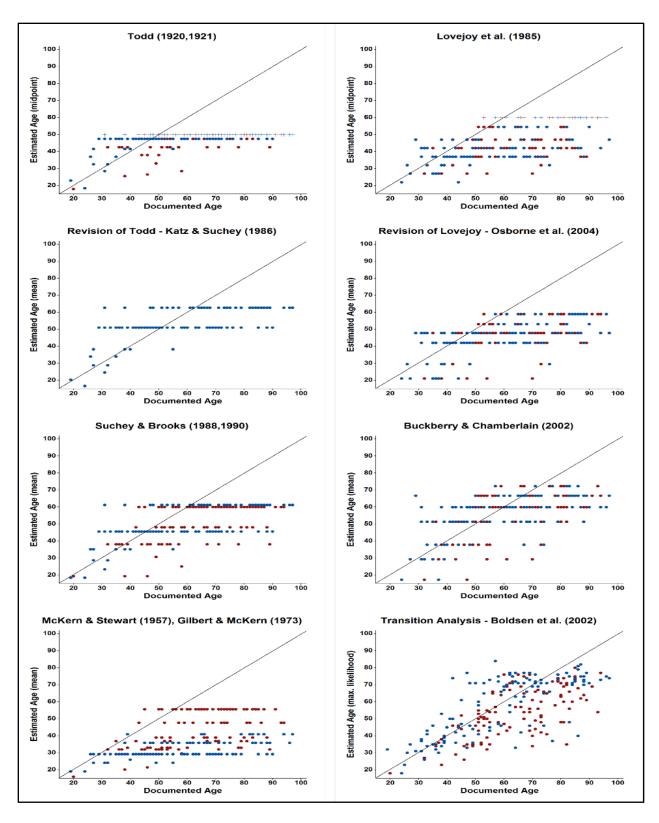


Figure 4.0. Age estimates using standard procedures for the sacroiliac joint and pubic symphysis and existing TA based on both the pelvis and cranium. Dots indicate published central tendencies and plus signs indicate the lower bounds of terminal open-ended intervals (males in blue and females in red). Figure created by Getz with Milner data (Bass Collection, N=234); data previously presented in (Milner, Boldsen, Ousley, Weise, et al., 2016).

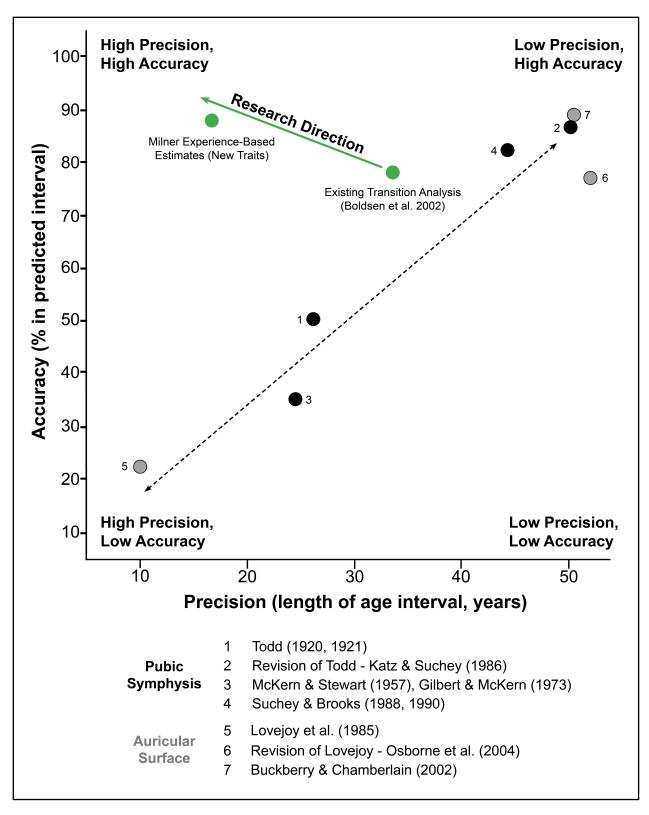


Figure 4.1. Existing methods produce estimates along the line from low accuracy with high precision (lower left) to high accuracy with low precision (upper right). Figure created by Getz with Milner data (Bass Collection, N=234); presented in (Milner, Boldsen, Ousley, Weise, et al., 2016).

NIJ Data Collection

Skeletal samples

The NIJ project relies primarily on data from four modern known-age skeletal collections: the William M. Bass Donated Collection (Forensic Anthropology Center, 2012), the Pretoria University Collection (L'Abbé, Loots, & Meiring, 2005), the Chiang Mai University Collection (Mahakkanukrauh, Khanpetch, Prasitwattanseree, & Case, 2013), and the Bocage Museum Collection (H. F. Cardoso, 2006). Several identified forensic cases and donated individuals from Mercyhurst University were also evaluated. Table 4.0 lists the location and origin of each skeletal collection and the primary ancestry of the individuals evaluated.

Table 4.0. Skeletal collections evaluated by the NIJ project in Phase 2

Skeletal Collection	Location	Collection Origin	Primary Ancestry ¹
WM Bass Donated	Knoxville, TN (USA)	donated/forensic	American white and black
University of Pretoria	Pretoria, South Africa	donated	South African white and black
Chiang Mai University	Chiang Mai, Thailand	donated/forensic	Thai
Bocage Museum	Lisbon, Portugal	cemetery	European white
Mercyhurst University	Erie, PA (USA)	donated/forensic	American white

¹ Individuals from other ancestry groups were evaluated at the collections when available, but sample sizes are insufficient for separate analyses.

Identifying and refining skeletal traits

Prior to the first round of NIJ-funded data collection in May 2015, the 45 features identified in preliminary work by Getz in Phase 1 (see Table 3.3) were combined with 30 additional traits suggested by members of the research team. Nine features of the pubic symphsis and auricular surface used in existing TA were also added with a reduced number of categories. All of the features are easily visible with the naked eye and no specialized equipment or destructive sampling is necessary for data collection.

All features were assessed by six members of the research team in modern skeletal samples representing populations from North American, Europe, Africa, and Asia (Table 4.0). After each round of data collection and analysis, existing trait definitions were refined by experienced osteologists, both native and non-native English-speakers, with different levels of familiarity with the features. Definitions were revised, diagrams were added, and trait variants and scoring exceptions were extensively documented through notes and photographs.

This process is a slightly elaborated version of the process used in Phase 1 of this dissertation (see Figure 3.3). After each round of data collection, the data were entered into electronic form and preliminary analyses were conducted by Getz using R code collaboratively developed by the research team. These results were initially discussed by Milner and Getz and the results, suggestions for trait eliminations and modifications, and points of discussion were disseminated to the research team. Additional input was gathered from team members in the field regarding trait modifications and was incorporated into the scoring procedures by Getz in revisions of the trait manual after each round of data collection.

More advanced analyses of the NIJ data set, primary conducted by Boldsen, are currently underway. Features that show consistent age-related patterns between the sexes and among populations will be incorporated into the new age-estimation procedure. Table 4.1 shows the samples where reference data were collected, the number of traits and transitions identified, and the features eliminated after each round of analysis. During data collection in Pretoria and Chiang Mai, several features were added and a number were modified. Therefore, to have as complete a data set as possible, Milner and Getz returned to the Bass Collection in February 2016 to obtain these data.

Table 4.1. Summary of Phase 2 trait refinement and data collection process

		Ske	letal Sai	nple	Traits			
Collection	Date	N	М	F	New	Used	Eliminated	Transitions
Bass	05/2015	423	221	202		89	15	129
Pretoria	07/2015	424	269	155	3	77	1	107
Chiang Mai	01/2016	440	271	169	3	79	0	118
Bass(2)	02/2016	14	12	2	0	79	0	118
Mercyhurst	04/2016	7	5	2	0	79	0	118
Lisbon	06/2016	390	190	200	0	79		118
Final		1698	968	730				

In addition to identifying new features throughout the skeleton, modifications have been made to the existing TA features in the cranium and pelvis. Categories with transition curves that span many decades and contribute little to overall age estimates have been eliminated, which resulted in a more streamlined scoring process. The number of trait variants for each feature was further reduced based on discussions among the members of the research team. The remaining stages were redefined to make them easier to score and less susceptible to intraand inter-observer error. Carefully worded descriptions were elaborated with diagrams and high-quality images with informative captions primarily developed by Milner and Getz.

Establishing trait age distributions

As briefly discussed in Chapter 3, the existing version of TA uses a generalized linear model (the logit) to estimate the probabilities of traits occurring at each age across the adult lifespan (Boldsen et al., 2002). Generalized linear models are similar to the more commonly used linear regression models that many individuals are familiar with from introductory statistics. Simple linear regression models involve a single continuous predictor variable (x), a univariate response variable (y), some unknown parameter to be estimated from the data (b), and a constant (a) (Equation 4.0). This model assumes a linear relationship between the predictor and the response variables and interpretations of the model assume that the response variable has a normal distribution (S. N. Wood, 2006e). Generalized linear models allow the expected value of the response variable to depend on a smooth, monotonic function of the predictor variable and the distribution of the response variable can follow any distribution from the exponential family (S. N. Wood, 2006c). In the case of TA, it is the natural log of the odds of a trait being present (logit of p) that is linearly related to the product of the predictor variable (age, x) and estimated coefficient (β_1) , plus the constant (β_0) (Equation 4.1). The response variable (p) is assumed to have a binominal distribution. Maximum likelihood estimation is used to find the best fitting values for the constant and coefficient.

Once the constant and coefficient in the model have been estimated, the inverse logit (or the inverse of Equation 4.1) must be used find the probability of a trait being present for an individual of each age. For this, the logit link (Equation 4.2) is used. Applying Equation 4.2 to the linear predictor (the right side of Equation 4.1), the probability of a trait being present in the jth skeleton $y_i = 1$ with age-at-death (x_i) is shown in Equation 4.3 (Boldsen et al., 2002). To obtain an age estimate in the existing Transition Analysis program, the probabilities are combined to produce a maximum likelihood age-at-death estimate in a manner discussed later in this chapter. The basic steps between Equations 4.1 and 4.3 are shown in Appendix C.

$$y = a + bx + \varepsilon \tag{4.0}$$

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \varepsilon \tag{4.1}$$

$$logit link = \Lambda = \frac{e(\cdot)}{1+e(\cdot)}$$
 (4.2)

$$\Pr(y_{j} = 1 \mid x_{j}) = \Lambda \left(\beta_{0} + \beta_{1} x_{j}\right) = \frac{e^{(\beta_{0} + \beta_{1} x_{j})}}{1 + e^{(\beta_{0} + \beta_{1} x_{j})}}$$
(4.3)

One alternative under investigation for future versions of TA is the substitution of smoothed empirically based curves in place of those currently produced by logistic regression. These alternative curves are produced with generalized additive models (GAMs) fitted using the R package mgcv (R Core Team, 2012; Wood, 2004; Wood, 2006). Generalized additive models are a type of generalized linear model where the linear predictor involves a sum of smooth functions of covariates (S. N. Wood, 2006d). In essence, rather than assuming that the entire reference data set can be described by a distribution with a single parameter value, generalized additive models allow for more flexibility by fitting multiple parameter values to different parts of the data set. 14

In the GAMs used here, the complexity of the curves – how much 'wiggliness' is allowed – is controlled by a smoothing parameter value (λ). This parameter value can be automatically chosen by the gam() function through repeated iterations of model fitting and generalized cross validation (GCV)¹⁶ or specified by hand (S. N. Wood, 2006a, 2006d). A smoothing parameter value of 0 results in an unpenalized model with as much 'wiggliness' as is justified by the data, while a large smoothing term will produce a completely smooth curve that essentially duplicates the logistic function. If the smoothing parameter is too low, the function will be under-smoothed and fit the underlying signal as well as the noise in the data. The model will perform poorly when applied to other data sets. In contrast, if the parameter is too high, the curve will be oversmoothed and potentially obscure important features of the data. The optimal smoothing parameter is a compromise between smoothness and the fit of the model (S. N. Wood, 2006d).

¹⁴ Whether GAM models fitted using regression splines should be considered parametric or non-parametric is debatable because "once the number of knots is chosen, a parametric family has been specified with a finite number of parameters" (Faraway, 2006b, p. 219).

Term used by (S. N. Wood, 2006b).

¹⁶ GCV can be turned on by adding (k=-1) to the model formula. Wood (2006a, 2006d) provides a more detailed discussion of generalized cross validation and the specifications of generalized additive models.

The research team is currently investigating three possible models for incorporation into the new Transition Analysis framework, the logit, a GAM with an automatically chosen smoothing parameter, and a GAM with a smoothing parameter hand-selected from a range of ten values ranging from 0.000001 (essentially unpenalized) to 1000 (essentially logistic). Figure 4.2 compares the curves generated for three features using the smallest (λ =10⁻⁶) and largest (λ =10³) of the smoothing parameters. It is likely that the optimal smoothing parameter for many features, the best compromise between model complexity and practical value, lies somewhere between the extremes.

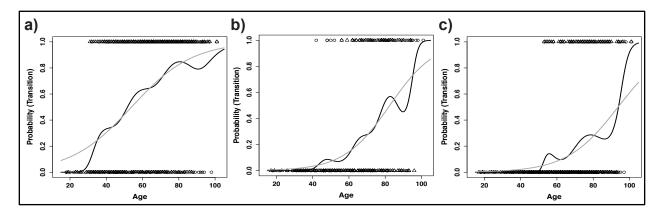


Figure 4.2. Comparison of the GAMs generated with the smallest (10⁻⁶, black line) and largest (10³, grey line) smoothing parameters for a) lateral clavicle macroporosity, b) sternal dorsal ridges, and c) thoracic vertebrae candlewax. Data from Getz Phase 1 primary reference sample.

Preliminary investigations by members of the research team (Milner pilot data) indicate that GAMs may show steeper ages of transition than those produced by logistic regression for some traits. Because steeper transitions provide tighter age intervals for individual features, it is possible that using trait probabilities from GAMs may improve the precision of overall estimates generated from many traits. Additionally, in some cases, the added complexity in the generalized additive model may represent real and, ultimately, important underlying features of the data. However, there is also the risk in small samples that the features of a more complex model may just as easily represent noise. As an added complication, although GAMs may more accurately reflect reality, common tasks, such as hypothesis testing and the estimation of intervals, are computationally more intensive (S. N. Wood, 2006b). Therefore, the potential improvement to age estimates provided by generalized additive models must be balanced against the time, complexity, and uncertainty introduced into the process.

Figure 4.3 compares the logistic and GAM models with an automatically selected smoothing parameter generated for three traits. The divergence between the models is relatively small, but shows the greatest difference at the ends of the age distribution. These small differences may, cumulatively, affect the final age estimate produced, but the impact of model choice will likely vary depending on the age of the individual being assessed and the particular mix of characteristics present.

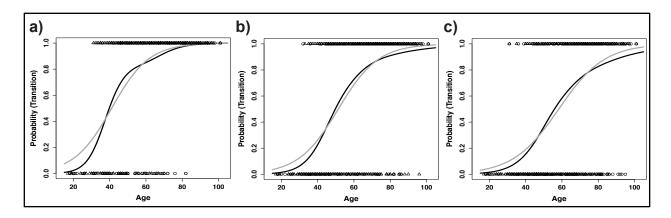


Figure 4.3. Comparison of the logistic (grey line) and generalized additive model (black line) for the a) profile of ribs three through ten, b) humerus lateral epicondyle, and c) lumbar vertebral lipping, with the greatest divergence occurring at the ends of the age distribution. Data from Getz Phase 1 primary reference sample.

The difference between the models generated based on the hand-selected smoothing parameter and the logistic curves vary greatly by feature. In some cases, such as those shown in Figure 4.4, the choice of smoothing parameter has essentially no effect. The logistic curves and all ten generalized additive models [$\lambda = 10^{-6} \ 10^{-5}, \ 10^{-4}, \ 10^{-3}, \ 10^{-2}, \ 10^{-1}, \ 10^{0}, \ 10^{1}, \ 10^{2}, \ 10^{3}$] are, for practical purposes, the same.

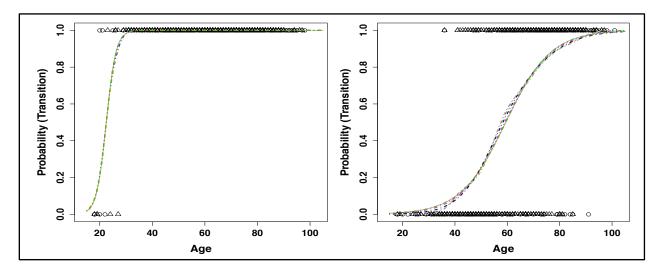


Figure 4.4. Two features for which the choice of smoothing parameter has essentially no effect. Logistical curve (solid grey line) and ten different generalized additive models with different smoothing parameters (colored dotted lines) for medial clavicle (left) and bone growth on the lateral margin of the lesser tubercle of the humerus (right). Data from Getz Phase 1 primary reference sample.

In some cases, such as those in Figure 4.5, the choice of smoothing parameter has a slight, but noticeable, effect on the shape of the fitted model, particularly at the ends of the age distribution. For curves in this category, the bumps in middle age likely reflect slight fluctuations in the sample age distribution, but are unlikely to relate to underlying biological processes.

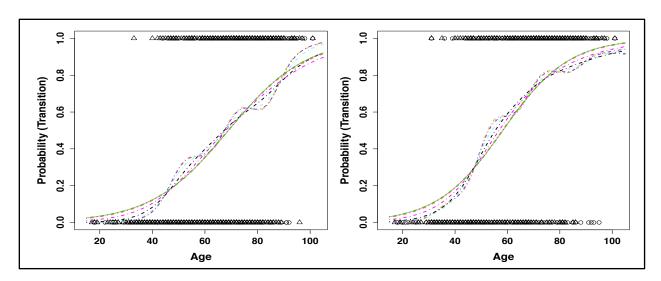


Figure 4.5. Two features where the choice of smoothing parameter produces slight differences. Bumps in middle age likely reflect slight fluctuations in the sample age distribution, but are unlikely to be related to underlying biological processes. Logistic curve (grey solid line) and ten GAMs (colored dotted lines) are shown for textural changes on the medial epicondyle of the humerus (left) and lipping of the lumbar vertebrae (right). Data from Getz Phase 1 primary reference sample.

For a small number of features, such as those in Figure 4.6, the choice of smoothing parameter makes a significant difference in the age-informative value of the feature. The models for these traits fall into one of two groups: 1) significant variability throughout life with a rapid increase in old age, or 2) a steady increase throughout adulthood with, ultimately, only a small fraction of the population at each age having the feature. Although the features of the more complicated models may prove to provide interesting glimpses of underlying biological processes, for the purposes of age estimation, the more conservative, smoother curve is likely to be of more use for age estimation.

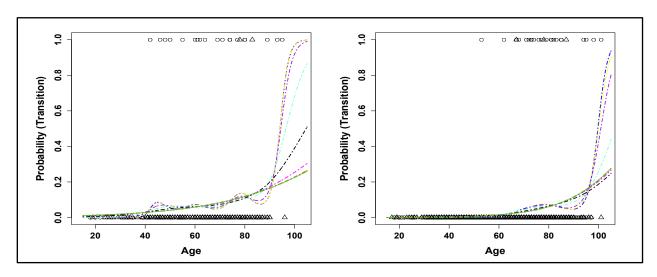


Figure 4.6. Examples of traits where the smoothing parameter has a significant effect on the shape of the fitted curves, especially at the oldest ages. Models shown for hyperostosis frontalis interna (HFI) (left) and extremely thin ribs (right). Data from Getz Phase 1 primary reference sample.

Selective mortality

Another issue currently under investigation by the research team is the possible influence of selective mortality on trait distributions. It is possible that for some traits, individuals who develop the feature may have an increased risk of death relative to their age-matched counterparts who do not. Although selective mortality cannot be detected with smoothed logistic functions, it may be possible to detect this signal using generalized additive models. For some features, such as those shown in Figure 4.7, the logistic curve (grey line) shows a general increase from young to old age, while the GAM with an automatically selected smoothing parameter (black line) shows variability in the latter half of the lifespan. The bumps around 60 years of age may possibly indicate selective mortality and are smoothed over by the logistic model. It is possible that in these cases the early development of the feature is correlated with some underlying disease or process that increases the risk of death. Once these individuals are removed, the trait frequency in the remainder of the population increases with age. It is possible that this phenomenon is partially responsible for the large amount of variation present in age estimates for middle-aged individuals, and the correspondingly large age ranges produced.

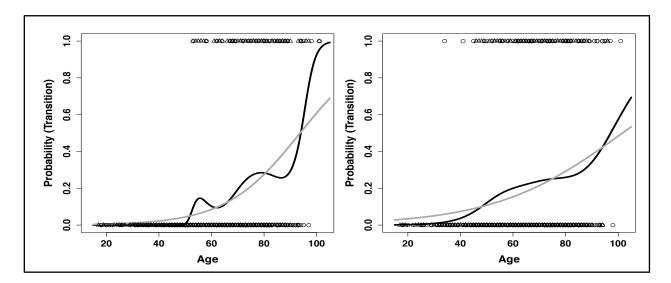


Figure 4.7. Logistic (grey line) and generalized additive models with an automatically chosen smoothing parameter (black line) for candlewax (DISH) in the thoracic vertebrae (left) and irregular ossifications on the profile of ribs three through ten (right). Data from Getz Phase 1 primary reference sample.

Similarly, some features show a pattern where the trait appears in the sample, increases in frequency with age, and the drops out relatively early in the age distribution (Figure 4.8). Although the association with selective mortality is speculative at this point, two of the features in Figures 4.7 and 4.8, thoracic candlewax and superior-anterior sacroiliac joint fusion, are associated with a condition known as diffuse idiopathic skeletal hyperostosis (DISH). This condition has been associated with adult-onset diabetes, obesity, hypertension, vascular and purine metabolism disorders, and the potential to cause severe spinal cord injury from relatively minor trauma (Hannallah et al., 2007; Kiss et al., 2002; Poelstra, 2013), so an increased risk of death, particularly for males, would not be surprising.

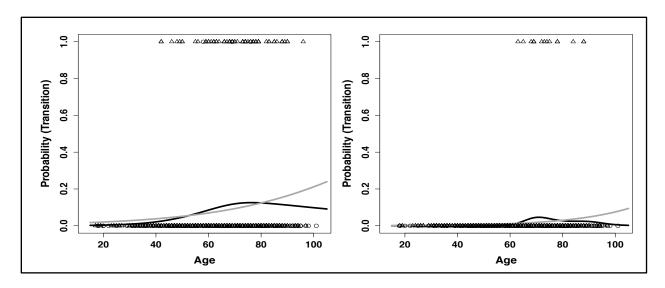


Figure 4.8. Logistic (grey line) and generalized additive models with an automatically chosen smoothing parameter (black line) for superior-anterior sacro-iliac joint fusion (left) and development of the pubic symphyseal collar (right). The bumps around 60 years of age may possibly indicate selective mortality and are smoothed over by the logistic model. Data from Getz Phase 1 primary reference sample.

Because of the relatively small number of individuals available from skeletal samples at the upper ends of the lifespan, it is possible that variation in the models can be explained by random chance in sampling. However, if similar model features also appear in the large and diverse sample collected by the research team, the possibly of selective mortality becomes more likely. Identifying these trends in skeletal samples is the first step towards more detailed investigations with larger samples of documented individuals and collaboration with medical researchers, which will be briefly discussed in Chapter 7. For the purposes of the NIJ project, identifying features that are heavily influenced by selective morality and removing them consideration may result in a reduction in age-estimate error in middle age.

Age-At-Death Estimates Using Multiple Features

Maximum likelihood estimates

The models discussed above generate individual likelihood functions for the probability of a trait existing at each age of the adult lifespan. The probability for multiple traits is equal to the product of their individual likelihood functions (Equation 4.4). To mathematically simplify this computation, the natural log of the likelihood function (i.e, log-likelihood) is used (Equation 4.5). The steps between Equations 4.4. and 4.5 are shown in Appendix C.

$$L = \prod_{j=1}^{n} \left[Pr(y=1 \mid x) \right]^{y_j} \left[1 - Pr(y=1 \mid x) \right]^{1-y_j} \tag{4.4}$$

$$Ln L = \sum_{j=1}^{n} \{ y_j \ln[\Pr(y=1 \mid x)] + (1 - y_j) \ln[1 - \Pr(y=1 \mid x)] \}$$
(4.5)

Using Equation 4.5, the individual log-likelihood values for each trait are summed to produce a likelihood function for the entire skeleton. The peak of this distribution is the maximum likelihood value of age. Although the math of the generalized additive models is not the same, a similar process of summing the log-likelihood values for each feature is used.

Preliminary age estimates for Bass Collection skeletons

Figure 4.9 shows the age estimates produced using the existing TA procedure based only on cranial and pelvic traits (a), compared to those produced using new features throughout the skeleton collected by the NIJ team (b). Both analyses used sex-specific transition curves and a uniform prior distribution. The existing TA procedure includes an *ad hoc* correction for correlated features that is not included in the preliminary Transition Analysis with new traits shown here (Boldsen et al., 2002). However, the correction for trait correlations only influences the width of the age interval and not the maximum likelihood age estimate. Therefore, Figure 4.9 essentially compares the results of the same statistical procedure using different features. Although the error in age estimates still increases in middle age, the addition of new traits has greatly reduced the age-estimation error throughout the lifespan and almost eliminated the age-estimate bias seen in existing techniques. The effects of correlated features on the precision of age estimates will be discussed in greater detail in Chapter 5.

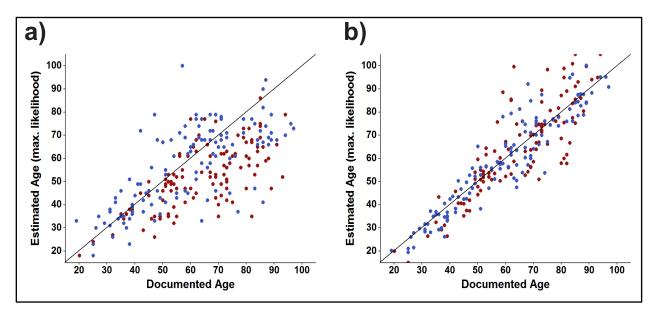


Figure 4.9. (a) Maximum likelihood age estimates produced using existing TA based only on cranial and pelvis traits, and (b) the new procedure using features throughout the skeleton for the same sample (males in blue and females in red). Identify line shown for comparison. Age estimates generated by Boldsen using Milner data (Bass Collection, N=234) and figure created by Getz; data presented in (Milner, Boldsen, Ousley, Getz, et al., 2016).

These preliminary TA estimates provide evidence that eliminating features with unexplained variation in middle age, potentially the result of selective mortality, may improve the maximum likelihood age estimates. Figure 4.10 compares new TA estimates generated for Bass skeletons using a small sample of traits including those showing possible selective mortality signatures, and estimates for the same set of skeletons with those features removed. In the first set of estimates (a), GAMs with hand-selected smoothing parameters were used to

generate maximum likelihood ages at death. In the second set (b), traits that showed variation in the sex-specific GAM curves were eliminated. Although estimates are still wider in middle age than at the ends of the age distribution, eliminating features based on the shape of the GAM curves seems to improve the correlation between known and estimated age. The impacts of trait and model selection on the accuracy and precision of age estimates will be further explored using two additional samples in Chapter 5.

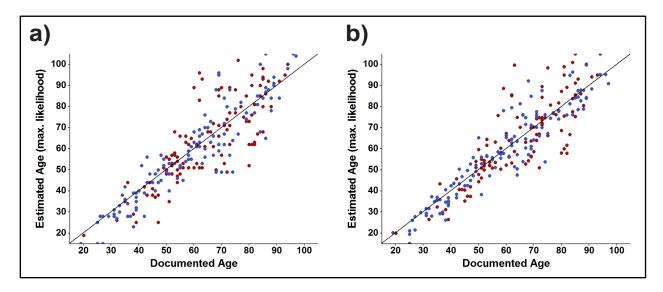


Figure 4.10. Comparison of new TA maximum likelihood estimates (males in blue and females in red) for the same sample based on (a) all features showing age-related patterns and (b) the same feature set with traits showing variation in the sex-specific GAM curves eliminated. Identify line shown for comparison. Age estimates generated by Boldsen using Milner data (Bass Collection, N=234) and figure created by Getz.

Effect of Model Choice on Maximum Likelihood Age Estimates

Maximum likelihood age-at-death estimates were generated by Getz for individuals from the Athens Collection (N=201) using Phase 1 reference data for 40 features using the same procedure as shown above. These estimates indicate that the choice of model, logistic or either of the GAMs makes relatively little difference in the point estimates of age. The pattern is the same whether a single model is used to represent the entire sample of males and females collectively (Figure 4.11), or if sex-specific models are used to estimates the ages of males and females separately (Figure 4.12). The effect of model selection when using different groups of features, as well as on the length of associated age ranges will be investigated further in the validation study in Phase 3 and by the research team in their larger sample.

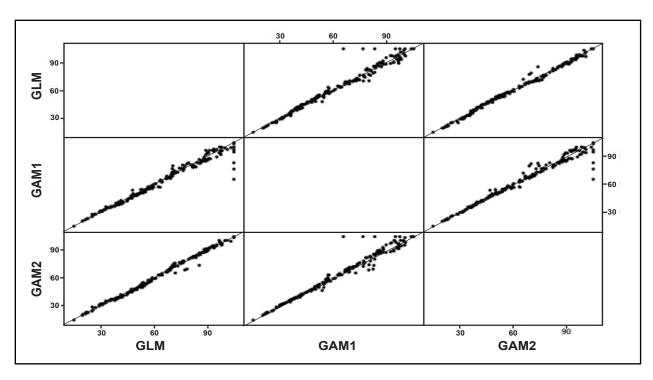


Figure 4.11. Comparison of maximum likelihood age estimates produced for the same individuals using a logistic model (GML), a generalized additive model with an automatically selected smoothing parameter (GAM1), and an alternative GAM with a hand-selected smoothing parameter (GAM2), all based on a combined sex reference sample. Identity line shown for reference. Getz data, Athens Collection, N=201.

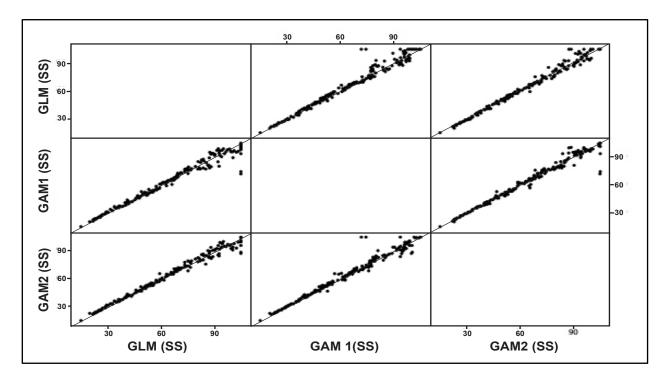


Figure 4.12. Comparison of the maximum likelihood age estimates produced for the same individuals using sex-specific models. Identity line shown for reference. Getz data, Athens Collection, N=201.

Age intervals for maximum likelihood estimates

In preliminary transition analysis trials by the research team, and those in this dissertation (Chapter 5), age intervals around the maximum likelihood estimate of age are estimated using the likelihood-ratio test. This test is based on the concept of forming a confidence interval around the single estimated parameter in our model (θ) , such that all of the values of (θ) for which the log-likelihood $l(\theta_0;x)$ are within a given tolerance of the maximum value $l(\hat{\theta};x)$. If θ_0 is the true value of the parameter, then likelihood-ratio test statistic (Equation 4.6) is approximately distributed as a chi-squared distribution with one degree of freedom $(\chi)_1^2$ when the sample is large (Boldsen et al., 2002; D. R. Cox & Hinkley, 1979).

$$2\log\frac{L(\hat{\theta};x)}{L(\theta_0;x)} = 2\left[l(\hat{\theta};x) - l(\theta_0;x)\right] \tag{4.6}$$

The likelihood ratio tests the null hypothesis $(H_0: \theta = \theta_0)$ versus a two-sided alternative $(H_1: \theta \neq \theta_0)$. The null hypothesis is rejected at the α -level if the likelihood ratio test statistic is greater than the value of the chi-squared distribution at the $100(1-\alpha)^{th}$ percentile of the chi-squared distribution (Faraway, 2006a). For a 95% confidence interval (α =0.05) the value is 3.841 (Equation 4.7). In other words, the 95% interval includes all values where the likelihood function decreases by no more than 1.9205 units¹⁷ (Equation 4.8).

$$2\left[l(\hat{\theta};x) - l(\theta_0;x)\right] \le 3.841\tag{4.7}$$

$$l(\theta; x) \ge \left[l(\hat{\theta}; x) - 1.9205\right] \tag{4.8}$$

This concept is visualized in Figure 4.13 for a 95% confidence interval. A line is drawn on the log-likelihood function at the value 1.9205 units below the maximum likelihood value for age (horizontal line). The end values of the age interval (vertical lines) are the points where the probability function intersects with the horizontal line. For a 99% confidence interval, the horizontal line would be moved down, and the confidence interval would widen accordingly.

¹⁸ Although hypothesis testing for GAM models is only approximate (S. N. Wood, 2006b), age estimates in the preliminary NIJ work and this dissertation are calculated in the same manner for all log-likelihood age functions, regardless of whether logistic or GAM curves were used.

¹⁷ The value of the 99th percentile of a $(\chi)_1^2$ is 6.6.35, so a 99% interval would include all values where the log-likelihood function drops off by no more than 3.317 units.

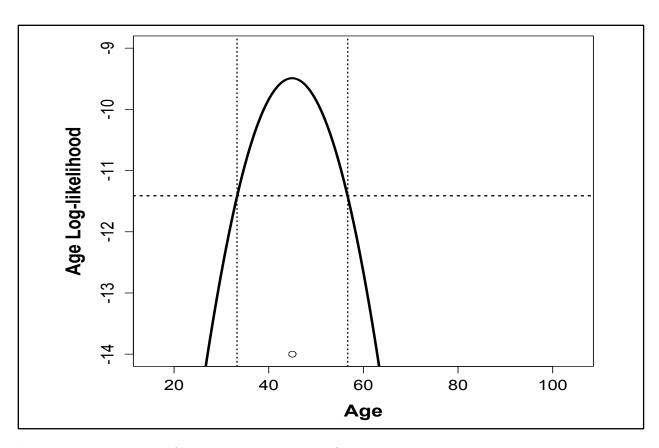


Figure 4.13. A log-likelihood function based on 33 skeletal features showing a) a horizontal dashed line 1.9208 units below the peak of the distribution and b) the 95% interval based on the intersection of the horizontal line with the probability curve (vertical dotted lines). Open circle represents the true age of a female from the Athens Collection.

Correlated traits

When combining features using a transition analysis approach, an assumption must be made that each trait provides independent information about age if the effect of age is removed. In other words, it is assumed that features are correlated with each other because they each independently develop with increasing age and that this effect can be mitigated statistically. Unfortunately, because underlying biological processes affect bone growth and resorption throughout the skeleton in similar ways and age-estimation methods often use multiple features from the same skeletal element or joint, this assumption is likely to be untrue.

If correlated traits are used to estimate age and an adequate statistical correction is not applied, then the age intervals produced will be systematically too narrow. This effect can be seen in tests of existing methods when lower accuracy is obtained than would be expected based on the use of confidence intervals of a particular length (e.g., 95%). This is precisely what has been found in tests of existing TA (e.g, Figure 4.1) despite the method's statistical correction for correlated features.

This issue is of particular importance for new TA procedure because of the large number of new features identified throughout the skeleton. The level of correlation among traits resulting from common underlying biological processes, between bilaterally scored features, and among traits in the same developmental or functional unit are currently unknown and potentially

complex. Because many traits provide similar, but distinct, information about age (e.g., the anterior superior margin of S1 and adjoining margins of L5), the decision of which features to include in the method will be based partially on the effect that including correlated traits has on the age estimation. Ideally, the number of features in the new TA will be enough to generate reasonably precise estimates, but not so many that the correlations among them result in overly narrow age estimates that reduce accuracy.

As part of the NIJ project, the research team is investigating statistical methods for mitigating the effects of correlated features as well as practical ways to identify and reduce the number of traits used. Correlations among trait pairs and groups of functionally and developmentally related features will be investigated using both NIJ and Getz dissertation data. The statistical methods are primarily the responsibility of other members of the research team, while several practical options have been evaluated in this dissertation. One of these solutions involves selecting the most age-informative feature from correlated pairs based on the slope of the transition curves and only the "best" one (i.e., the one with the steepest transition) is used to calculate an age-estimate. This procedure maximizes data from partial skeletons because age-related features are not removed prior to observation of the skeleton. In the new TA software, this practical solution may be implemented in such a way that traits are prioritized based on their age-informative value, and a less informative trait can only be entered if the preferred one is not available. Another solution involves iteratively searching for the set of features that minimizes age-estimate error, while maximizing accuracy. The results of these preliminary tests using the Athens collection are presented in Chapter 5.

The New Transition Analysis Method Software

Software development

Upon completion of NIJ project, anticipated for the end of 2017, the new TA method will be available as both a stand-alone program and as a part of Fordisc, an internationally used program used to statistically estimate the ancestry, sex, and stature of adult skeletal remains based on large reference samples (Jantz & Ousley, 2013; Ousley & Jantz, 2015). Steve Ousley, one of the PIs of the NIJ project and the developers of Fordisc, has already integrated the existing TA method into the newest version of this software, which should be available in early 2017. Fordisc is already routinely purchased by practitioners and academic institutions as part of their standard software used for research and training purposes. With the added capability of adult age-at-death estimation, it is likely to become the single most heavily used resource for generating biological profiles for medicolegal purposes. However, because Fordisc must be purchased, to reach the broadest possible audience, the new TA method will remain available as a free, stand-alone program.

To date, Ousley has already completed several updates to improve the functionality of existing TA in terms of compatibility with updated software platforms and the presentation and interpretation of results. Modifications to the way the program outputs results include the option to export results as a report and the addition of bar charts for estimated age ranges (Figure 4.14). Although the bar charts (Fig.4.14, right) are easier to interpret and explain to non-specialists, the likelihood curves (Fig. 4.14, left) provide an excellent visual representation of the variation present in the skeleton and better represent the contributions of each feature to the age estimate. Therefore, the program retains the capability to produce both types of output.

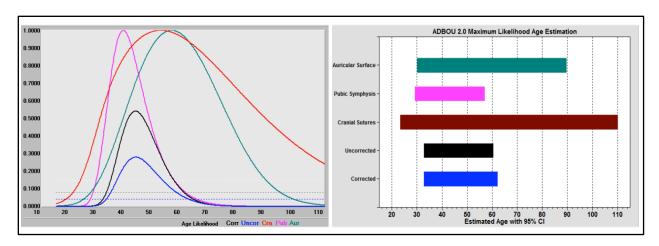


Figure 4.14. Probability curves and bar plot output from the existing TA program (ver. 2.1.041. The curves and bars show the age range produced by each of the areas of the skeleton individually (auricular surface—green, pubic symphysis—magenta, cranial sutures—red) and the combined age estimate using a uniform (black) or archaeological (blue) prior distribution.

At the completion of the NIJ project, Ousley will retain primary responsibility for updating both the free and Fordisc versions of the program as computer hardware and software change to ensure that they will not become obsolete. This ongoing support ensures that our data, incorporated within both program versions, will be available to researchers long after support for the NIJ project ends. Additionally, data collected using the new TA method from known-age skeletons around the world can be incorporated into the Forensic Data Bank (FDB), which has been managed by Ousley for over a decade. The FDB holds the reference data for Fordisc and also maintains data from over 3,400 modern individuals and many nineteenth century samples (Ousley & Jantz, 2012).

Sex and population-specific reference samples

The existing TA program allows users to use sex- and ancestry-specific reference data. The options for ancestry currently include black and white, which are the two groups represented in the Terry Collection originally used as the reference sample for the method (Boldsen et al., 2002). Selecting unknown for either sex or ancestry, results in the use of combined reference sample data. For example, if male with unknown ancestry is selected for analysis, probability models estimated from all black and white males in the reference sample will be used.

Based on preliminary analyses conducted by Getz, it is uncertain whether significant improvements will be seen in age estimates produced using sex- and ancestry-specific models for the new features because the curves for many features are similar. Figure 4.15 shows the maximum likelihood age estimates produced by Getz for individuals from the Athens Collection (N=201) using three different models fitted to both combined and sex-specific reference samples. These data demonstrate that the use of different models for males and females produces only minor differences between the point age estimates. Although differences in point estimates may be slight, they could potentially be significant in forensic contexts and greater differences may be seen in the associated age-interval lengths. Additionally, the differences between combined and sex- and population-specific models could vary significantly depending of the traits ultimately selected for analysis. These issues will be further discussed in the validation study in Chapter 5.

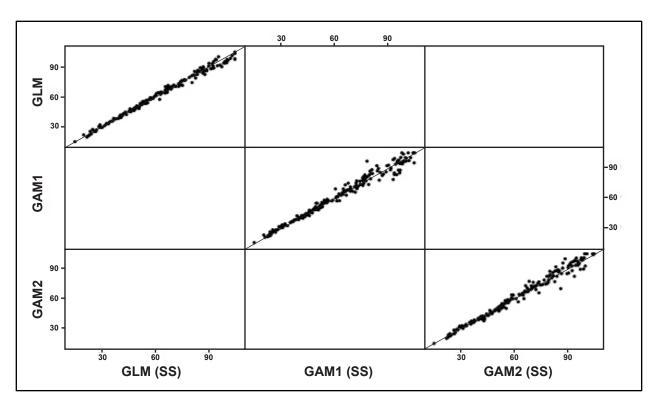


Figure 4.15. Maximum likelihood age estimates for the same individuals using three models generated from combined (left) and sex-specific (bottom) reference samples. Identity line shown for reference. Getz data, Athens Collection. N=201.

At present, the research team anticipates that the ability to select sex- and population-specific models will be retained in the new version of the TA program. Using the samples collected as part of the NIJ project, users will potentially have the option of white, black, Asian, and unknown. However, because of sample size constraints, it may be possible that the initial version of the new program may only have the option of white and unknown. The use of a combined reference sample will likely result in more conservative estimates (i.e., longer prediction intervals), but may be more appropriate for forensic use when ancestry is unknown or population-specific reference samples are unavailable.

Prior distributions

The simplified version of the TA procedure used to test the performance of the new features in this dissertation assumes a uniform prior distribution, which means that an individual of every age—15 to 105 years—is equally likely to die and end up in a mortality sample. Although a uniform prior distribution is a conservative approach because it does not impose a specific age-at-death structure onto the data, it is also an unrealistic one (Boldsen et al., 2002; Konigsberg & Frankenberg, 1994). Even in the United States where individuals routinely live beyond 50 years of age, the probability of finding a 90 year-old and 60 year-old in a mortality sample is not the same. Although the probability of death increases with age, there are also fewer individuals of each age in the population as a whole, so fewer of them will ultimately end up in mortality samples. Because a uniform prior does not account for the fact that all individuals with the oldest trait scores are unlikely to be at uppermost extreme of the age distribution, the ages of individuals who exhibit the oldest trait categories tend to be overestimated. Tests of existing TA

have shown that, in practice, a uniform prior does in fact overestimate age at older ages (Boldsen et al., 2002; Milner & Boldsen, 2012c).

The existing Transition Analysis program includes two options other than a uniform prior distribution. These are a "forensic" distribution, which includes all homicides compiled by the Centers for Disease Control and Prevention (CDC), and an "archaeological" distribution, which is derived from seventeenth century rural Danish parish records (Boldsen et al., 2002). The research team anticipates that several prior distribution options for medicolegal investigations may be added, including overall mortality, homicides, and forensic cases in the United States. The overall impact of these distributions on age estimates in both forensic and archaeological settings will also investigated by the research team.

Despite its shortcomings, in this dissertation a uniform prior distribution is used to assess the performance of the new features. Because this dissertation evaluates multiple samples that differ substantially in both known and unknown ways, a uniform prior distribution is the most likely way to obtain results that can be directly compared among the samples. This approach also simplifies comparisons between the existing TA method and the new procedure. Fortunately, it has been shown that the choice of prior distribution has little effect on maximum likelihood ages estimates until after age 50, and only a relatively small effect until after age 80. Figure 4.16 compares existing TA estimates generated using uniform and archaeological prior distributions for skeletons from the Athens Collection. The pattern is nearly identical to what was found by Milner and Boldsen (2012) in a published validation study of existing TA using a different skeletal sample.

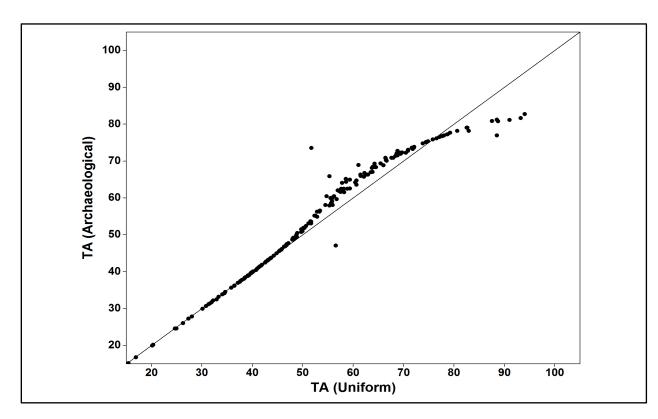


Figure 4.16. Comparison between maximum likelihood age estimates produced using the uniform and archaeological prior distributions in existing TA. The identify line is shown for reference. Getz Data: Athens Collection, N=201.

Trait scoring manual

The NIJ project began data collection in May 2015 at the Bass Collection with a trait manual consisting of 28 pages of text. During each data collection trip over the next year, members of the research team made suggestions, corrections, and notes in personal copies of the manual, which they were encouraged to refer to frequently. All comments and edits were compiled by Getz and key points were discussed as a group before modifications to trait definitions were made. As new traits were identified, definitions were developed and refined through this process of suggestion and discussion. Team members also flagged specific skeletal elements or individuals during the data collection process for photographic documentation. Photos were taken to represent both the obvious and subtle versions forms of a trait. Borderline cases, situations where features should not be scored, and common points of confusion were also documented and these images are discussed with informative captions in the trait manual.

Having non-native English speakers as part of the research team (Boldsen, Tarp, and Weise) helps to ensure that the terminology and descriptions used for the features are intelligible to researchers worldwide. Although Ousley is primarily responsible for the computer programming aspect of this project, he has also made important contributions to improving trait definitions because he was not initially involved in their development. Standardizing terminology and simplifying definitions is an ongoing process and will likely continue even after the first version of the new TA manual is released. The final version of the trait manual will also include a glossary, which clearly identifies and defines key terms used throughout the text.

The scoring manual used for the final scheduled round of data collection in Lisbon, Portugal consisted of 86 pages with 26 diagrams and 440 images. Additional diagrams illustrating important concepts, such as scoring areas and trait locations, are currently under development and several dozen additional images from the last round of data collection are awaiting integration into a new draft of the manual. Versions of manual figures and images will also eventually be incorporated into the computer software for the method. Sections of the NIJ scoring manual describing the traits used to test the new TA procedure in Chapter 3 are included as Appendix D.

CHAPTER 5: PHASE 3—TESTING REVISED TRANSITION ANALYSIS ON TWO KNOWN-AGE SAMPLES

An ideal age-at-death estimation method reliably produces accurate and precise age estimates with a quantifiable degree of certainty and is equally practical for individual skeletons and large samples (Milner & Boldsen, 2012b). A reliable method consistently produces the same results regardless of the observer or the number of times that it is applied. Precision is related to the length of the estimated age range for a particular individual; a technique with high precision provides a narrow age interval, while low precision results in broad ranges. In this dissertation, accuracy is used to describe whether the documented age of an individual falls into the estimated range. 19 Ideally, age intervals generated by a method should be wide enough to capture the true age of an individual, while still being narrow enough to provide useful information (Buckberry, 2015). Unfortunately, many existing methods obtain reasonably high levels of accuracy by decreasing precision to the point where at least some intervals cover most of the adult lifespan (e.g., Brooks & Suchey, 1990; DiGangi et al., 2009; Meindl & Lovejoy, 1985; Osborne, Simmons, & Nawrocki, 2004). To be practical, the method should be based on parts of the skeleton that are often recovered and typically well preserved in forensic and archaeological contexts. The process for analyzing features should be quick, non-destructive and require no specialized equipment or extensive training to apply reliably. To have confidence in the results of a method, the technique's applicability to different populations across time and space, and the likely consequences if the recommendations are violated, should be known.

Although they may appear to be similar, the criteria for a useful age estimation method in archaeological settings are somewhat different that those for forensic purposes. Age estimation in forensic settings typically involves the analysis of a single individual or a very small number of them. Estimated sex, ancestry, and other contextual information can often be used to select a method that is likely to provide the most accurate and precise result for each case. For archaeological applications, however, researchers typically deal with larger samples of much less well-preserved remains. Critical elements, or important pieces of them, are often missing and taphonomic changes can hinder the interpretation of age-related changes. It is also often unclear what effects of applying methods developed using individuals from different populations are on archaeological samples. This issue of whether time period- and population-specific methods are required to adequately estimate adult age-at-death has become a contentious topic in recent years (Hoppa, 2000; Konigsberg et al., 2008; Molleson, 1995; Ubelaker, 2008).

The majority of anthropologists seem to believe, quite reasonably, that the use of a method developed on a group more similar to the population of interest will provide better results (A. Schmitt et al., 2002). Unfortunately, the number of well-documented skeletal reference collections in the world is far fewer than the number of populations that forensic anthropologists are interested in. Additionally, these samples are typically small, and often biased, samples of the populations they supposedly represent (Komar & Grivas, 2008; Usher, 2002). This means that any population-specific standards developed are based on a relatively small number of individuals who may or may not be a good match to the population as a whole or to the individual being analyzed (Milner & Boldsen, 2012b). Assessments of differences between populations using traditional methods typically apply the technique to a local, known-age skeletal sample or to two different groups (e.g., Gocha et al., 2015; Moraitis, Zorba, Eliopoulos, & Fox, 2014; Rissech et al., 2012; Sakaue, 2006; A. Schmitt, 2004) and the average age of

¹⁹ Accuracy is sometimes also used in published studies to describe how closely the estimated point age approximates an individual's documented age.

individuals assigned to each stage or, in rare cases, the age-at-transition between stages, are compared. Even relatively minor differences are often deemed evidence that population-specific standards should be used and revised age ranges for the new group are published. Although regionally-specific techniques have recently been developed for a number of populations, they are rarely used by most anthropologists (Falys & Lewis, 2011; Garvin & Passalacqua, 2012). This is potentially because interpreting the results of these methods is more complex, particularly the majority of the comparative data for the population or region are based on age estimates from standard techniques (Falys & Lewis, 2011).

Although the differences found between populations may be real, and potentially interesting, it is also possible that they are the result of deviations in the application of the method, the structure of the test sample, or random variations introduced by the particular individuals analyzed (Wärmländer & Sholts, 2011). For these reasons, among others, Konigsberg and colleagues (2008) emphasize the need to collect and analyze age-related changes from large samples, rather than focusing on inter-population differences and the development of a multitude of new regionally specific standards.

Phase 3: Method Validation

In this chapter, the performance of standard age-estimation methods and new TA are evaluated using two additional known-age skeletal samples. Both samples contain individuals of European ancestry, but represent different populations than those in the Phase 1 reference sample. The first—the Athens Collection—is used to assess how the method performs on modern, well-preserved skeletons. The skeletal features identified in Phase 1 (Chapter 3) are combined using the simplified TA procedure described in Phase 2 (Chapter 4). Seventy-two variations of the new TA procedure using different numbers of traits, statistical models, reference samples, and prediction intervals are tested to identify the strengths and weaknesses introduced by each factor. These tests are evaluated based on the overall accuracy, precision, and collective bias of the age estimates produced. The results are compared with the performance of the most commonly used age-estimation methods (Garvin & Passalacqua, 2012) to evaluate which procedures are the most effective for forensic and archaeological applications.

The best-performing new TA method variations identified using the Athens sample are then applied to individuals from St. Bride's crypt. This application assesses how the new TA method performs on less well-preserved individuals with significantly different diet and disease exposure than the Phase 1 reference sample. Because the historical sample is also significantly less well preserved, this test evaluates the performance of the new TA method on incomplete and fragmented remains, similar to those frequently encountered in archaeological contexts. Importantly, the individuals from St. Bride's crypt lived contemporaneously with those from one of the Phase 4 archaeological samples. Thus, the results of this test are used as an indication of the accuracy, precision, and bias that can be expected in this archaeological application. Ultimately, these results indicate whether the new TA procedure is capable of producing reliable results in an archaeological context. Information about the individuals evaluated from the two collections in Phase 3 is summarized in Table 5.0 and Figure 5.0.

Table 5.0. Skeletal collections used to test the new TA procedure in Phase 3

Collection	Location	Origin	Death Years	N	М	F
Athens Collection	Athens, GR	cemetery	1960 - 1997	201	111	90
St. Bride's Crypt	London, UK	crypt	1740 - 1852	168	92	76

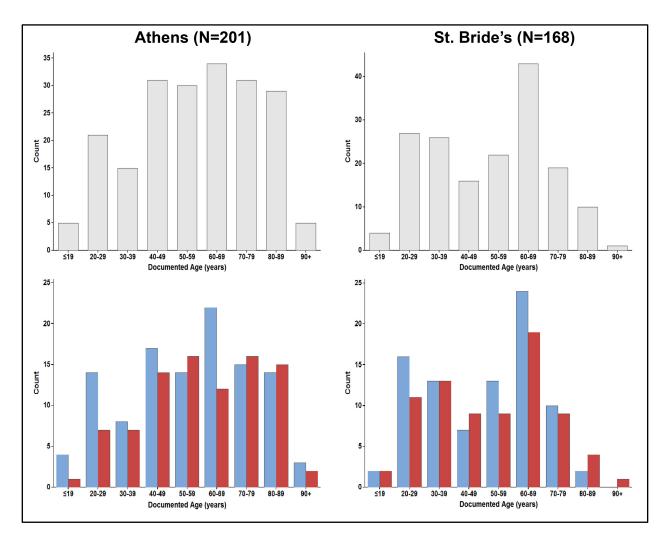


Figure 5.0. Age distributions of individuals evaluated from the Athens and St. Bride's samples (top row) and the same individuals separated by sex (bottom row, males in blue and females in red).

Phase 3 samples

Athens Collection

The University of Athens Human Skeletal Reference Collection, also known simply as the Athens Collection, was formed in two parts. The first 72 individuals were curated between 1996 and 1997 at the Wiener Laboratory of the American School of Classical Studies in Athens. In 1998, the collection was donated to the Department of Animal and Human Physiology at the

University of Athens and an additional 153 individuals were acquired between 2001 and 2003 (Eliopoulos et al., 2007). Age, sex, cause of death, place of birth, and occupation were obtained from death certificates and are available for 214 individuals. Individuals were collectively born in all areas of Greece and were interred in Athens or nearby areas between 1960 and 1997 (Eliopoulos et al., 2007). These individuals are most similar to individuals in death cohort 2 (post-1981) from Phase 1 and are used to investigate the effects of using probabilities estimated from a time period-specific reference sample to estimate age.

St. Bride's Crypt

In the past 1,000 years, eight churches have existed on the current site of St. Bride's Church in London, England. The seventh of these, famously designed by Sir Christopher Wren, opened in 1675 (St. Bride's Church) and, beginning in 1740, individuals were interred in the church crypt. Each individual was placed in a lead coffin that included a plate inscribed with the name, sex, age-at-death, and death date of the person inside. Burials continued until 1854 when all crypts within the city of London were permanently closed in response to public health concerns stemming from a severe cholera epidemic (Scheuer & Bowman, 1995). After the crypt was bricked over, the church stood for almost another century until it was left almost entirely destroyed during the Blitz in December of 1940 (Harvey, 1968; St. Bride's Church). In 1952, during a mandatory architectural survey conducted prior to the construction of the church that now stands on the site, the long-forgotten crypt was discovered (Scheuer & Bowman, 1995; St. Bride's Church).

Although little information exists on these excavations, it is known that individuals in the crypt were identified based on coffin plates before being placed into metal munitions boxes for transfer to Cambridge for analysis (Scheuer & Bowman, 1995). In some cases, multiple individuals, usually from the same family crypt, were placed into a single box for transport. Therefore, the possibility exists that remains may have become commingled; this likely affects only a small number of individuals, and discrepancies are noted in the collections database (J. Bekvalac, Museum of London, personal communication, Nov. 4, 2015). After analysis, individuals were transferred into plastic boxes with their associated coffin plates and, beginning in 1990, were transferred again to their present storage configuration (Scheuer & Bowman, 1995). Skulls, postcranial remains, and coffin plates are stored separately in the basement (crypt) of St. Bride's Church (Scheuer & Bowman, 1995).

For some of the individuals who died after 1837, information from coffin plates and parish registers was used to locate death certificates, which provided additional data, such as place and cause of death (Scheuer & Bowman, 1995). This documentary evidence suggests that individuals buried in the crypt were, on average, of higher socioeconomic status that those buried in the cemeteries outside of the church (Scheuer & Bowman, 1995). However, the assessment of socioeconomic status is relatively subjective, because it relies heavily on stated occupation, which tended to change frequently during this time period and is sometimes difficult to interpret (Harvey, 1968). Regardless of the exact details, on the whole, individuals in this sample are more similar to those in death cohort 1 (pre-1953) in terms of diet and disease exposure, than to individuals who died in the late twentieth century. The effect of generating age using probabilities from both a combined and death cohort-specific reference sample will be assessed using this collection.

Establishing a baseline for method improvement

As was discussed in Chapter 4 in the context of the NIJ project, to assess the performance of the new TA method, a baseline must be established for the performance of commonly used techniques. Individuals from the Athens Collection (N=201) are used to evaluate standard methods for the pubic symphysis (Brooks & Suchey, 1990; Katz & Suchey, 1986; Konigsberg et al., 2008; Suchey, Brooks, & Katz, 1988; Todd, 1920, 1921), auricular surface (Buckberry & Chamberlain, 2002; Lovejoy, Meindl, Pryzbeck, et al., 1985; Osborne et al., 2004), and existing TA (Boldsen et al., 2002), as well as variations of the new TA procedure. These methods were selected because they are the most-commonly used procedures by forensic anthropologists and archaeologists (Garvin & Passalacqua, 2012). In the interest of time, cranial sutures were not evaluated other than in the context of existing TA.²⁰

Figure 5.1 compares the results of commonly used methods applied by Getz to those produced by Milner using the same procedures on a different sample. Histograms (a,b), show the age distribution of individuals in each of the test samples; scatterplots (c,d) compare the accuracy (percent of cases where the documented age falls within the estimated range) and precision (average age interval length in years) of the methods tested. Although the methods were applied by observers with differing levels of experience to different samples, the pattern produced is almost identical. The only significant deviation between observers is in experience-based estimates, which is to be expected given Milner's decades of experience working with skeletons from around the world.

All standard methods are either relatively accurate or precise, but one is sacrificed at the cost of the other. ²² In other words, these techniques provide either narrow intervals (high precision) that rarely contain the known age of the individual (low accuracy), or give extremely wide age ranges (low precision) that are often correct (high accuracy). Neither option is particularly useful in either forensic or archaeological contexts. The tests by both Milner and Getz show that experience-based estimates provide the best compromise between accuracy and precision, with existing TA performing the best of the standard methods evaluated.

In this chapter, the statistical framework of existing TA is applied to a broader array of ageinformative traits, especially those providing information in middle to old age. The ultimate goal is to produce accurate, precise, and unbiased age estimates that approximate those of experienced observers, but that come with a quantifiable degree of certainty and a procedure that can be easily taught to others.

_

Over 100 years ago, the American anatomist Thomas Dwight (1890, p.389) discussed the poor age information provided by cranial sutures: "It is, I believe, pretty generally admitted among anatomists that the time and order of the closing of the cranial sutures are very uncertain, far too much so for the them to be trustworthy guidelines to determine the age of the skull." A century of research has failed to alter the truth of this statement (M. Cox, 2000; Hershkovitz et al., 1997; Milner & Boldsen, 2012c). There is little reason to belabor this point here.

Hershkovitz et al., 1997; Milner & Boldsen, 2012c). There is little reason to belabor this point here.

21 Two additional procedures (Berg, 2008; Katz & Suchey, 1986) were evaluated but are not included in the figure because ages cannot be estimates for both males and females using the published standards. Therefore, the accuracy and precision of their estimates are not directly comparable to those of the other methods.

accuracy and precision of their estimates are not directly comparable to those of the other methods. ²² Todd (1920,1921) originally described ten non-overlapping phases, with ranges of one to five years. The relatively high accuracy and precision shown here is an artifact of the final age range (Phase 10, 50+) being used in the calculations of average age interval length as 50 to 100 years. Below age 50, the method is highly precise with low accuracy. Above 50, the method is highly accurate with extremely low precision.

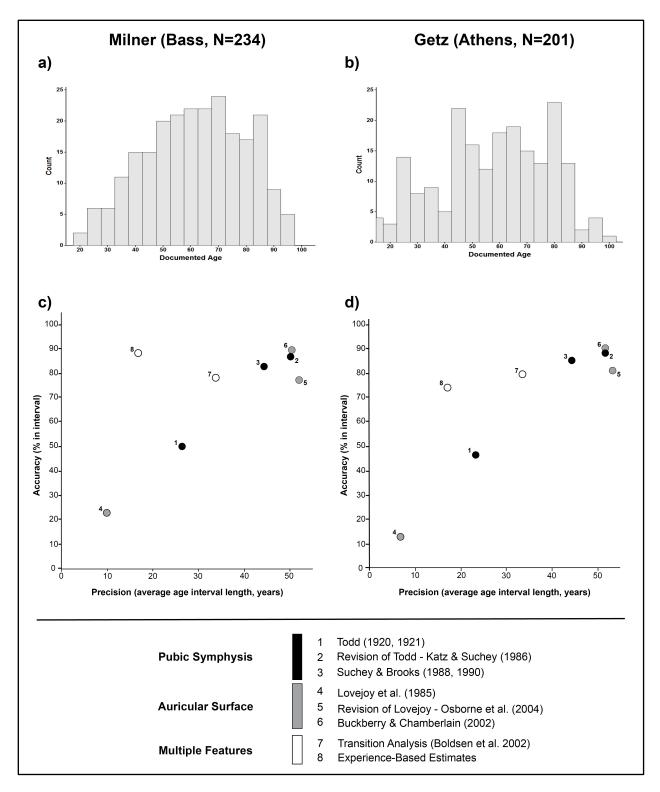


Figure 5.1. Age estimates generated by Milner (c) and Getz (d) for different samples (a,b) using commonly used pubic symphysis (black circles) and sacroiliac joint (grey circles) methods, existing TA, and experience-based estimates (white circles) showing the same pattern for different samples.

The New TA Procedure

In Phase 1, 1,010 individuals were evaluated from four well-documented North American skeletal collections—WM Bass Donated, Maxwell Museum Donated, JCB Grant, and Ul-Stanford. Using this combined sample, the age-informative potential of fifty-three features was evaluated using both logistic and generalized additive models (Chapter 3). Forty-five features that showed relatively consistent patterns in each of the collections were selected for further testing by the research team (Chapter 4 and Appendix B). During Phase 2, the NIJ research team added and revised the definitions of many features, including substantial revisions to five Phase 1 traits. Because the data for these features were not necessary consistent with how they were collected in the Phase 1 reference sample, they were eliminated from further consideration in this dissertation. As a result, the variations of the new TA procedure evaluated here include a maximum of 40 features. The portions of current draft version of the NIJ scoring manual describing these traits are provided in Appendix D.

In the new TA procedure, estimating the age of an unknown skeleton using these features is a two-step process. First, probability curves are generated for individual traits using a skeletal reference sample. Second, the individual probability functions for each trait are combined based on the features and their character states observed in a skeleton to produce a maximum likelihood age-at-death estimate and range.

Establishing probability density functions for skeletal features

To generate the probability density function for a skeletal feature three decisions must be made: 1) the statistical model to be used; 2) the sample of reference data that the model will be fitted to; and 3) whether a single model will be fitted to the combined sample or if sexspecific models will be generated.

As described in Chapter 4, the NIJ research team is currently investigating three slightly different models for generating probability functions from skeletal features: 1) a generalized linear model, the logit (GLM); 2) a generalized additive model with an automatically selected best-fit smoothing parameter (GAM1); and 3) a generalized additive model with a hand-selected smoothing parameter (GAM2). Although, in some cases, the added complexity of the generalized additive model may represent real and, ultimately, important underlying features of the data, there is also the risk that in small samples more complex models features may just as easily represent noise (S. N. Wood, 2006b). Additionally, fitting GAMs is more computationally intense and the statistical theory related to testing the fit of the model, generating confidence intervals, and comparing alternative models is less well understood (S. N. Wood, 2006b). For these reasons, the improvement in the accuracy and precision of estimates based on generalized additive models must be balanced against the time, complexity, and uncertainty introduced into the process.

_

²³ Hyperostosis frontalis internal (HFI), the ossification pattern of the first rib, lipping of the occipital condyles, ossification along the ventral edge of the pubic symphysis (symphyseal collar), and spur development on the superior margin of the ischial tuberosity were eliminated from consideration in this dissertation, but remain under investigation by the NLI research team.

by the NIJ research team.

24 The majority of these features (75%) originated with Milner, with the remainder suggested by Getz during Phase 1 data collection. Milner and Getz were jointly responsible for writing new trait definitions and refining these after each round of dissertation data collection. Definitions were further clarified and improved based on suggestions by all members of the research team after each round of NIJ data collection. All pictures were taken by Milner or Getz and incorporated into the manual in its current draft form by Getz, with input from all members of the NIJ team.

Once a model is selected, it is applied to a sample of reference data to generate age-specific probabilities for the presence of a skeletal feature. The Phase 1 reference sample (N=1010) can be analyzed as a whole or divided into two "death cohorts": 1) pre-1953—individuals who died between 1924 and 1952 and 2) post-1981—individuals who died between 1982 and 2013. This division roughly separates the sample into groups that, on the whole, lived with substantively different diet, activity level, disease exposure, and medical care. All individuals in the chosen reference sample, either all of the individuals from Phase 1 or one of the death cohorts, can be fitted with a single model for all individuals or divided by sex.

Figure 5.2 shows the eighteen combinations of the three model types (GLM, GAM1, and GAM2), reference sample (all Phase 1 individuals or death cohort 2), and combined or sex-specific divisions that can be evaluated. Unfortunately, there are too few females in death cohort 1 to allow sex-specific probability curves to be generated, so only 15 different variations are assessed in this dissertation.

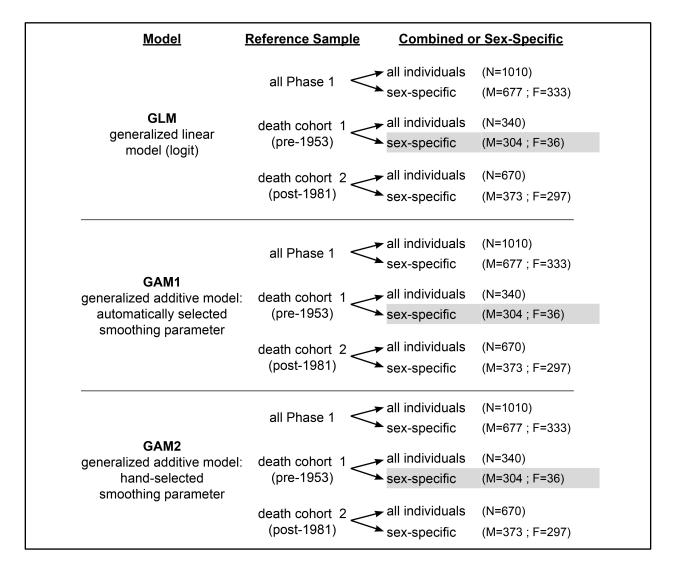


Figure 5.2. Eighteen variations of how probabilities can be calculated for traits from the reference sample collected in Phase 1. Only fifteen are evaluated here because there are too few females in death cohort 1 (grey boxes) to allow sex-specific probability curves to be generated.

Combining features to estimate age

To generate a maximum likelihood age-at-death estimate and associated prediction interval for an individual, the probability functions for each trait observed in the skeleton are combined. As illustrated in Figure 5.2, the probabilities for each trait can be generated in one of eighteen ways based on slightly different statistical models and reference samples. To simplify calculations, the natural log of each trait probability (i.e., log-likelihood) is used to calculate a cumulative log-likelihood function for each individual. Equation 5.0 shows the log-likelihood function for an individual with n skeletal traits. Each trait in an individual (y_i) can have a value of 0 (absent) or 1 (present), which is what determines which probabilities are added to the likelihood function. If the trait is present, then y_i = 1, the second half Equation 5.0 becomes 0, and only the probability of the trait being present is added to the likelihood function. If the trait is absent, then y_i = 0, the first half of Equation 5.0 becomes 0, and only the probability of the trait being absent is added.

$$Ln L = \sum_{i=1}^{n} \{ y_i \ln[Pr(y=1|x)] + (1-y_i) \ln[1 - Pr(y=1|x)] \}$$
 (5.0)

As described in Chapter 4, a 95 or 99% prediction interval around the maximum likelihood age estimate can be calculated using a procedure based on the likelihood-ratio test. The test is based on the concept of using a hypothesis test to form a confidence interval around the single parameter estimated in the logistic model (Boldsen et al., 2002; D. R. Cox & Hinkley, 1979). Although this procedure is widely used for logistic models, the statistical basis for its use on the other models tested here is somewhat less justified because hypothesis testing for GAMs is only approximate (S. N. Wood, 2006b). However, this procedure is currently used to estimate prediction intervals for all of the likelihood functions, regardless of the model used to generate them, to facilitate comparisons between them. If the decision is made to incorporate the generalized additive models into the new TA procedure, other members of the NIJ research team will further investigate whether this procedure is appropriate and potential alternatives.

Correlations between traits

As also discussed in Chapter 4, probabilistically combining features in a transition analysis framework requires the assumption that each trait provides independent information about age (Boldsen et al., 2002). If correlated traits are used to estimate age then the intervals produced will be systematically too narrow. Although existing TA includes a statistical correction for correlated features (Boldsen et al., 2002), the accuracy of the method is less than would be expected given the use of 95% prediction intervals (see Figure 5.1). This likely indicates that the effects of trait correlations have not been entirely removed using the current correction.

This issue is of particular importance for new TA procedure because of the large number of new features identified throughout the skeleton. The correlations between and among them are currently unknown and potentially complex. Ultimately, the new TA method must balance the need to use enough traits to produce reasonably precise estimates with the potential reduction in accuracy resulting from trait correlation. No statistical correction for trait correlations is currently included in the tests of new TA procedure presented here. So, as a preliminary assessment of the relationships between trait pairs, the Phase 1 reference data set (N=1010) was used to produce a correlation matrix. The correlation coefficients can be used a rough gauge of the probable impact of correlated traits in the new TA procedure. A coefficient of one means perfect agreement, while zero indicates no relationship. A positive coefficient indicates that when one trait is present the other is also likely to occur, with a higher number indicating a greater probability of the traits both appearing in the same individual. A negative coefficient indicates the opposite relationship, where if one trait is present, the other is less likely to occur.

The full correlation matrix for all trait pairs is provided in Appendix E. The extent to which the magnitude of each correlation affects the estimated age interval and the level of correlation between features that will ultimately be deemed acceptable is currently unknown. Therefore, the correlation coefficients in the appendix are colored by the strength of the relationship (yellow: 0.3–0.499, orange: 0.5–0.599, red: 0.6–0.99) to aid in the assessment of patterns. For the sake of simplicity, p-values were not included in this preliminary assessment, but will need to be evaluated in future analyses. More sophisticated methods for assessing associations between traits pairs and groups of features will be investigated as part of the larger research team effort.

At first glance, the correlation matrix primarily reveals relationships that knowledge of the skeleton suggests should occur. For example, the four bone "weight" scores—subjective evaluations of the relative heaviness of the bones for their size²⁵— exhibit some of the highest correlations among all of the trait pairs assessed (Table 5.1). Because the factors resulting in lower than average bone density (i.e., osteopenia) tend to have effects throughout the skeleton rather than targeting specific areas (Mundy, 1999), it should not be surprising that bone weight scores are correlated both with increasing age, as well as with each other. Also, not surprisingly, features located on the same skeletal element, such as those found on the lesser tubercle, medial epicondyle and lateral epicondyle of the humerus (Table 5.2) show high correlations.

Table 5.1 Correlation matrix for the subjective weights of four bones

	calcaneus	humerus	os coxa	tibia
calcaneus		0.482	0.541	0.627
humerus	0.482		0.625	0.709
os coxa	0.541	0.625		0.608
tibia	0.627	0.709	0.608	

Table 5.2 Correlation matrix for four traits on the humerus

	latEpicondyle	medEpicondyle	lesserTubBumps	lesserTubLip
latEpicondyle		0.519	0.349	0.520
medEpicondyle	0.519		0.388	0.505
lesserTubBumps	0.349	0.388		0.541
lesserTubLip	0.520	0.505	0.541	

However, not all trait correlations are as intuitive. For example, "shingle ribs"—flat, brittle ribs, with extremely thin cortical bone—and depressions of the parietal bones, which typically associated with thinning of the skull, were assumed by the research team to be related to osteoporotic processes. Surprisingly, although these features may have some relationship with light bones elsewhere in the skeleton, the traits were not highly correlated.

2

²⁵ As part of the NIJ project, these four skeletal elements are now measured and weighed to generate a rough, but objective, measure of bone "density." Weight is also scored subjectively by research team members so the correlation between subjective and objective bone weight scores can be assessed. Length and weight data were collected for the Athens and St. Bride's skeletons, but cannot be used here because reference data from Phase 1 are not available.

Tables 5.3 shows that while pairs of bones weight scores for the calcaneus, os coxa, humerus and tibia are highly correlated (between 0.482 and 0.709), these traits have much lower correlations with thinned parietal bones and shingle ribs than would be expected if all six features were strongly influenced by the same underlying factors.

Table 5.3 Correlation matrix for traits based on thin and light bones

	calcaneus wt.	os coxa wt.	humerus wt.	tibia wt.	parietal depression	rib 3-10 thickness
calcaneus wt.		0.541	0.482	0.627	0.089	0.201
os coxa wt.	0.541		0.625	0.608	0.126	0.212
humerus wt.	0.482	0.625		0.709	0.120	0.193
tibia wt.	0.627	0.608	0.709		0.125	0.207
parietal depression	0.089	0.126	0.120	0.125		0.143
rib 3-10 thickness	0.201	0.212	0.193	0.207	0.143	

The relationship between lipping and candlewax in different areas of the spine (Table 5.4) is also not as clear as might be expected. Lipping in different portions of the spine is fairly strongly correlated with lipping in each of the other areas of the spine, but the same relationship is not seen with the presence of candlewax. To further complicate the issue, lipping and candlewax do not appear to be highly correlated in the cervical and lumbar vertebrae, but have a moderate correlation (0.40) in the thoracic region (Table 5.4).

Table 5.4 Correlation matrix for lipping and candlewax in three areas of the vertebral column

	cervLip	thorLip	lumLip	cervWax	thorWax	lumWax
cervLip		0.393	0.480	0.133	0.268	0.103
thorLip	0.393		0.503	0.091	0.402	0.132
lumLip	0.480	0.503		0.132	0.320	0.139
cervWax	0.133	0.091	0.132		0.188	0.218
thorWax	0.268	0.402	0.320	0.188		0.302
lumWax	0.103	0.132	0.139	0.218	0.302	

Only a few negative correlations were found and they suggest potentially interesting biological relationships. Table 5.5 shows several of the negative correlations that were found between traits involving excess formation of bone (cervical candlewax and superior-anterior fusion of the sacroiliac joint) and bone loss (weight of the innominate and shingle ribs). A negative relationship between these traits fits with the typical pattern that individuals tend to be "bone formers" or "bone losers" but not both; the factors governing this relationship are poorly understood (A. Schmitt, Wapler, Couallier, & Cunha, 2007).

Table 5.5 Correlation matrix for several traits based on excessive bone formation or resorption

	cervWax	SI Fusion	os coxa wt.	rib 3-10 thickness
cervWax		0.130	-0.018	-0.026
SI Fusion	0.130		-0.010	-0.047
os coxa wt.	-0.018	-0.010		0.212
rib 3-10 thickness	-0.026	-0.047	0.212	

Evaluating Variations of the New TA Procedure: The Athens Sample

Age estimates were generated for individuals in the Athens sample using either all 40 of the features identified in Phase 1 or a sub-group of 30 features selected by the NIJ research team. Probabilities generated using twelve of the variations in Figure 5.2 were combined to produce maximum likelihood point estimates and ranges using both 95% and 99% prediction intervals. These tests resulted in forty-eight²⁶ sets of ages and ranges that were evaluated for overall accuracy and precision. Collective bias is evaluated later in this chapter.

Figure 5.3 shows the average accuracy and precision of the ages generated using different method variations for the same sample. As expected, using a wider prediction interval results in wider ranges with correspondingly higher accuracy (95% PI—filled symbols, 99% PI—open symbols). On average, a 99% versus and 95% prediction interval increases the average age interval width by 6.3 years and increases accuracy by around 13%. In the Athens sample, this increase translates to an average of 26 additional individuals who were accurately aged (known age fell into the estimated range) using a wider prediction interval.

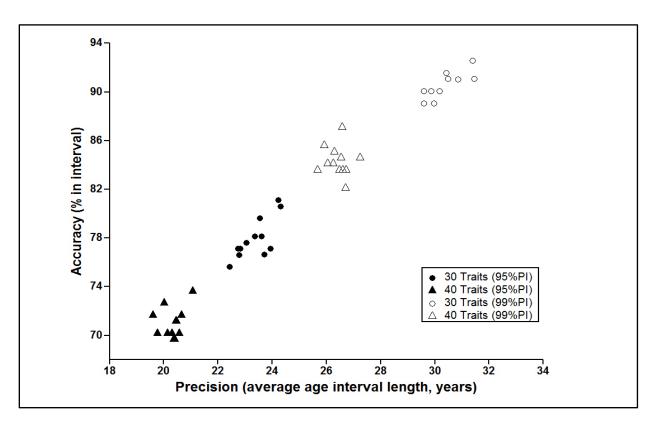


Figure 5.3. Each of the four point clusters shows the average accuracy and precision of age estimates for the entire Athens sample using trait probabilities from 12 slightly different model variations (shown in Figure 5.2). Differences between clusters are the result of the number of features (circles/triangles) and the prediction interval (filled/open) used to calculate the maximum likelihood estimate and its associated range.

²⁶ These tests do not include probabilities generated using death cohort 1 (pre-1953) because it would be an inappropriate reference sample for the modern individuals of the Athens Collections. The effect of using an inappropriate reference sample will be considered later in this chapter.

²⁷ The same 12 model variations were tested for each combination of traits and prediction interval; however, not all 12 symbols are visible in each cluster because some variations produced the same average accuracy and precision.

Comparing the estimates generated using all 40 features (triangles) to those based on a subset of 30 features (circles) reveals the same pattern of decreased precision yielding increased accuracy. Although this finding is not surprising based on the way accuracy is defined, it is important in these preliminary evaluations to verify that the method is performing as expected. It is also worth noting that in Figure 5.3, estimates generated using 30 features (circles) have higher accuracy than those based on 40 features (5.3) for all method variations. Although perhaps counterintuitive, more data do not necessarily result in better age estimates. As previously discussed, the use of correlated traits reduces the overall accuracy because of an inappropriate narrowing of the prediction intervals. This effect can be seen in Figure 5.3 as the obtained accuracy in all tests is lower than what would be expected based on the use either a 95 or 99% prediction interval.

Figure 5.4 shows the results of the same tests as in Figure 5.3, but groups the results by the type of model (GLM, GAM1, or GAM2) and reference sample (combined or death cohort 2), as well as whether combined or sex-specific probabilities were used. Within each cluster, models fitted to the combined reference sample (black symbols) generally showed the best compromise between accuracy and precision. Neither the logistic nor the generalized additive model consistently generated better age estimates.

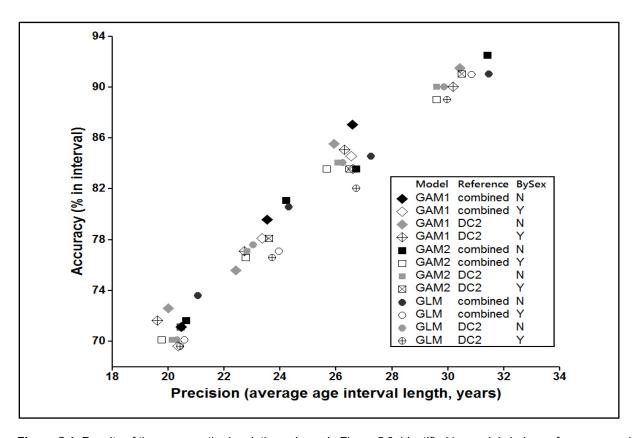


Figure 5.4. Results of the same method variations shown in Figure 5.3, identified by model choice, reference sample, and whether or not sex-specific probabilities were used. Variations using a combined, non-sex-specific reference sample resulted in the best compromise between accuracy and precision. No other consistent patterns are present.

76

²⁸ A similar phenomenon has been discussed by Stephen Ousley in the context of the appropriate number of traits to use in a discriminant function analysis (DFA) to estimate ancestry (Ousley & Jantz, 2012).

Although conventional wisdom suggests that models generated using samples of individuals who are more similar to the individual being evaluated would generate more accurate and precise results, this was not seen here. This is possible because the logistic regression and the GAMs used here require large samples to effectively characterize information. Therefore, the variation introduced by a reduction in sample size for each of the subgroups could be responsible for the less accurate age estimates seen in the ages generated with time periodand sex-specific models.

Using a reduced feature set to estimate age

Estimating age with a reduced number of features (30 versus 40) results in a slightly better compromise between accuracy and precision. To investigate whether other reduced-size feature sets would further improve age estimates, two approaches for selecting alternative groups of traits were evaluated. The first is a subjective approach where features are hand-selected based on their perceived age-informative value based on experience and the consistency of the information obtained from different model variations. The second approach is a two-step iterative search process designed to identify the smallest number of features capable of providing adequate coverage of the adult lifespan. To simplify interpretation of the results, both procedures are evaluated using only probabilities generated by logistic models

Hand-Selected Features

A subset of all possible features was selected in a two-step process based on expert judgment. First, the transition plots for the logistic and ten generalized additive model curves with different smoothing parameters fitted to the combined reference sample were reviewed for all forty features. Twenty-three features that showed relatively little variation among the models were identified. In step two, highly correlated trait pairs were identified using the correlation matrix in Appendix E. The transition curves of the features in each pair were re-evaluated and only the most age-informative feature was included in the final trait set. Care was also taken to ensure that the traits selected provided adequate coverage of the entire lifespan. Fourteen features were ultimately selected for testing.

Iterative Search

The first step in the process is to identify four traits that collectively provide good coverage of the entire adult lifespan to "seed" the iterative search algorithm used in step two. ²⁹ There are 91,390 possible combinations of the forty traits and each set can produce a maximum of sixteen maximum likelihood age estimates. The "best" combination of four features is the one that minimizes the absolute mean error of the age estimates produced for the entire test sample. In essence, the best combination is one where the sixteen possible age estimates fall closest to the greatest number of skeletons in the test sample. Because of this, if this search procedure is used on a test sample with an irregular age distribution, it will simply pick the combination of features that best mimics that structure.

_

²⁹ A set of four traits was chosen to seed the algorithm because it provided a sufficient number of possible maximum likelihood age estimates to adequately cover the adult lifespan while remaining computationally practical. A search for the best five features (32 possible maximum likelihood estimates) would require a program run time of approximately two and half days to evaluate all 658,008 combinations – likely for only a minimal improvement.

To avoid this scenario, a training set of 300 individuals with an essentially uniform age distribution was selected from the Phase 1 reference data. Age was estimated for all 300 individuals in the set using each of the possible combinations of four features. Because the test sample had an essentially uniform age structure, the four features that were selected were those where the sixteen possible maximum likelihood ages at death provided the best coverage of the entire adult lifespan. Figure 5.5 shows the logistic transition curves generated from the full Phase 1 reference sample for the four features selected.

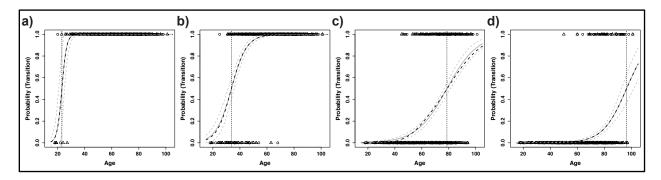


Figure 5.5 Four traits identified as the combination of features that provided the best coverage of the entire adult age range: a) medial clavicle epiphysis fusion, b) scapula glenoid fossa margin, c) humerus lesser tubercle bumps, and d) humerus weight.

In the second step, the four features and their associated absolute mean error are used as a starting point, or seed, for an iterative search algorithm. The program begins by adding the next feature on the list to the original four and estimating the age of all individuals in the test sample using the new group of five features. The process is repeated for each of the traits on the list and an absolute mean error is calculated for the sample based on the age estimates produced by each set of five features (the original four selected as a seed, plus a fifth from the list). The set of five features that produces the lowest mean error is then used to start the next round of evaluation, where each of the remaining traits is added to those already selected and the absolute mean error of the estimates produced by each group of six features is assessed. This process continues until adding another trait fails to improve the age estimates.

It is important to note that this is a greedy algorithm, meaning that each step of the program makes an immediately beneficial decision in the hope of arriving at an optimal solution with a minimal number of steps. In each iteration, the procedure selects the single trait that improves the absolute mean error of the estimates when used in combination with the group of traits already selected and continues. The algorithm does not evaluate every possible combination of traits in each loop. So, for example, rather than evaluating the entire test sample of 300 individuals using all possible combination of six features—3,838,380 possible combinations—the program only tests the performance of 36 groups, each containing the already selected set of five features plus one of remaining features on the list.

³⁰ Even with a reference sample of over 1,000 individuals from four skeletal collections to choose from, it was not possible to get enough individuals at the ends of the distribution (under 30 and over 90 years of age) to simultaneously have an equal number of individuals at each age and a sufficiently large test sample size. Three hundred individuals was the best compromise between obtaining a large enough test sample and maintaining an essentially uniform age distribution.

A set of 21 features was selected using this algorithm. Figure 5.6 compares the known age distribution of the "uniform" test sample with the estimated distribution obtained using the iteratively selected traits. The TA maximum likelihood estimates (b1,b2) do not exactly mimic the uniform distribution (a1,b1) when age is broken down into approximately one-year (a) or five-year (b) increments, but these data show two important features. First, when the estimated ages are grouped into five-year categories, the new TA procedure provides a relatively good approximation of the age distribution of the test sample. This is particularly heartening because these estimates were generated with the simplest of the statistical models tested here (logistic) with a combined reference sample, using non-sex-specific probabilities, and without specifying a particular prior distribution or correcting for correlated features. Second, the traits selected by the algorithm are capable of producing maximum likelihood age-at-death estimates at essentially every year of the adult lifespan between 15 and 105 years. Although the point age-at-death estimates shown here tend to under-represent the number of people over 80 years of age, the ranges typically capture the true age of individuals in this portion of the lifespan.

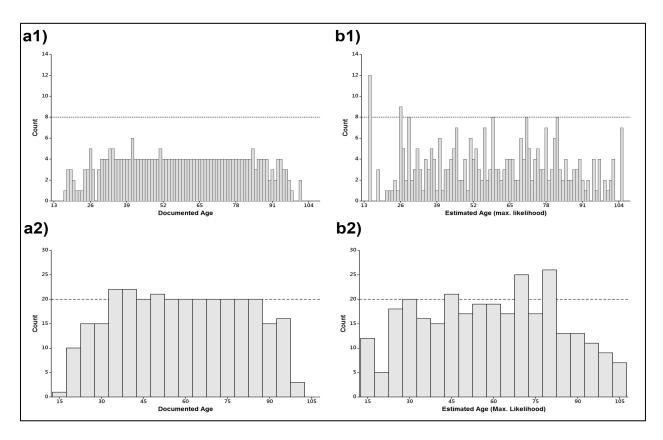


Figure 5.6. Age distributions of the known (a1,a2) and estimated ages of the "uniform" test sample (N=300) produced using the iteratively-selected features (b1,b2), shown in approximately one-year (a1,b1) and five-year (a2,b2) categories.

Table 5.6 lists all 40 features considered in this phase and identifies those included in each of the reduced feature sets evaluated. Figure 5.7 compares the average accuracy and precision obtained using 14, 21, or 30 features to that produced when using all 40 features for the individuals in the Athens sample. The results are shown for both a 95% (filled symbols) and 99% (open symbols) prediction intervals.

Table 5.6 Combinations of traits tested in variations of the new TA procedure

		Features Used					
	Trait	40	30	30 21			
1	C1 eburnation	Х	Х	Х	Х		
2	C1 lipping	X	X				
3	calcaneus weight	X	X				
4	cervical candlewax	X	X	X			
5	cervical lipping	X	X				
6	clavicle lateral macroporosity	X	X				
7	clavicle medial epiphysis fusion	X	X	X	X		
В	clavicle medial bone growth	X	X	X			
9	femoral fovea margin lipping	X	X	X	Х		
10	humerus lateral epicondyle	X	X	X			
11	humerus lesser tubercle bumps	X	X	X	X		
12	humerus lesser tubercle margin lipping	X	X				
13	humerus weight	X	X	X	X		
14	innominate weight	X	X	X			
15	ischial tuberosity bone growth	X	X	X	X		
16	L5 inferior margin lipping	X	X				
17	L5 superior margin lipping	X	X	X	X		
18	lumbar candlewax	X	X				
19	lumbar lipping	X	X		X		
20	medial trochanteric fossa exostoses	X	X				
21	acetabulum posterior margin lipping	X	X				
22	parietal depression	X	X	X			
23	R310 body thickness (shingle ribs)	X	X				
24	S1 superior margin (round)	X	X		X		
25	scapula glenoid fossa margins (round)	X	X	X	X		
26	thoracic candlewax	X	X				
27	thoracic lipping	X	X				
28	tibia weight	X	X				
29	trapezium lipping	X	X	X	Х		
30	trochanteric fossa exostoses	X	X	X			
31	AIIS exostoses	X		X	Х		
32	acetabulum articular surface bone growth	X					
33	femoral head surface bone growth	X					
34	humerus medial epicondyle	X					
35	R2 rim edge profile	X		X			
36	R310 rim edge profile	X					
37	radius medial crest	X		X	Х		
38	S1-2 fusion	X		X			
39	superior-anterior sacroiliac joint fusion	X		X			
40	sternal central dorsal ridges	X		X	X		

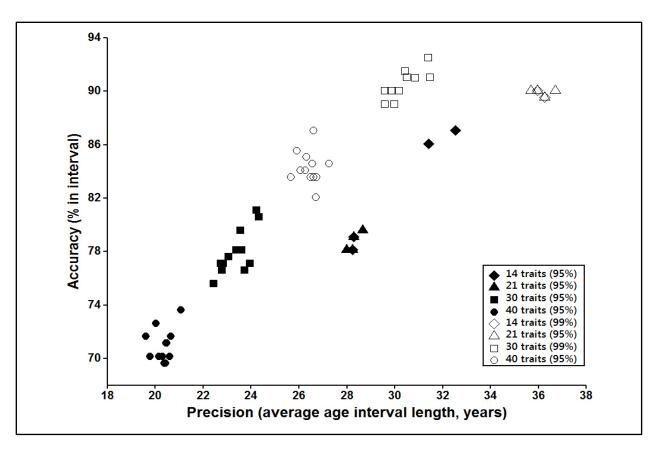


Figure 5.7. Comparison of new TA method variations using the logistic probabilities from different numbers of features. Circles (all 40 traits) and squares (30 traits) are the same data shown for both 95% and 99% prediction intervals. These are compared to the results using smaller samples of hand-selected (14, diamonds) and iteratively selected features (21,triangles). Athens Collection, N=201.

As expected based on previous tests of existing methods and the new TA procedure, using a wider confidence interval to estimate age increases the length of the age interval produced and increases accuracy. The two new sets of features—14 hand-selected traits (diamonds) and 21 interactively selected features (triangles)—show a slightly different relationship between their accuracy and precision. Surprisingly, the set of 21 features (filled triangles) have age ranges that are about five years wider (on average), but differ in average accuracy by less than 1% to ages produced using 30 features (filled squares). Although this is a relatively minor difference, it deviates slightly from what is expected. It is possible that this separation is an artifact of the relatively small sample tested. It is also possible that while ranges are increasing because there are fewer traits, there are also fewer correlated features that could be artificially narrowing the interval. The NIJ project will provide the larger data sets required to further investigate these patterns in different samples and subgroups of features.

Choice of reference sample: a cautionary note

In the variations of the TA procedure tested above, individuals from the Athens Collection were aged using probabilities from either a combined reference sample or only those in death cohort 2 (post-1981), which is likely to be the most similar to the Greek sample in many respects. These tests did not produce consistent improvements in the average accuracy and precision of age estimates using either time period or sex-specific reference samples. Figure 5.8 shows the

results of ages estimated using the same model, set of features, and prediction interval, but with different reference samples. Using a combined reference sample (black symbols) produces slightly better accuracy in all cases, with very similar estimates being produced using death cohort 2 (grey symbols). However, the use of death cohort 1 produces wider age ranges with reduced accuracy. In other words, the use of appropriate time period-specific probabilities did not result in improvement, but using an inappropriate reference sample appears to have a significant adverse effect.

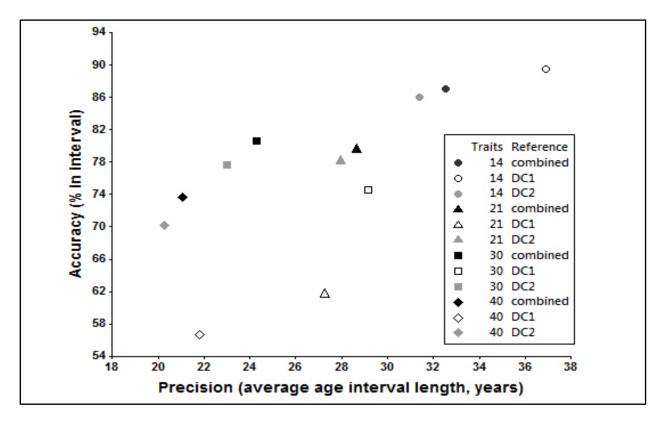


Figure 5.8. Comparison of ages generated with logistic model trait probabilities based on different sets of features and reference samples. Using a reference sample that is a poor fit for the test samples results in wider age ranges, lower precision, with reduced accuracy. Athens Collection, N=201.

The substantial decrease in accuracy obtained using death cohort 1 could be influenced by the smaller sample used to generate the trait probabilities. Because death cohort 1 has many fewer individuals at both ends of the age distribution than death cohort 2, the transition curves tend to be less consistent in those ages. Although the transition curves between the cohorts look similar, there is also the possibility that real and potentially meaningful differences in the skeletal aging process exist that cannot be detected using these samples.

This small test should be taken as a cautionary tale for the possible results of using methods based on inappropriate reference samples. In forensic contexts, and in essentially all archaeological ones, it is often impossible to identify which regional or population-specific reference sample would be the most appropriate. While using a diverse, combined reference sample could potentially decrease the precision of individual age estimates, the risks associated with choosing an inappropriate reference sample may be unacceptably high. For

most situations, the increase in applicability without sacrificing accuracy provided by using a combined sample is desirable. The NIJ project will be able to test whether the same result is seen using modern populations from around the world and contribute more substantive evidence to the population-specific method debate.

Performance of new TA relative to commonly used methods

Figure 5.9 presents the results obtained using the new TA procedure (95% prediction intervals only) alongside commonly used methods for the pubic symphysis and auricular surface, existing Transition Analysis, and experience-based estimates for the same individuals. Black symbols indicate variations of the new TA procedure, while standard methods, existing TA, and experience-based estimates are in grey.

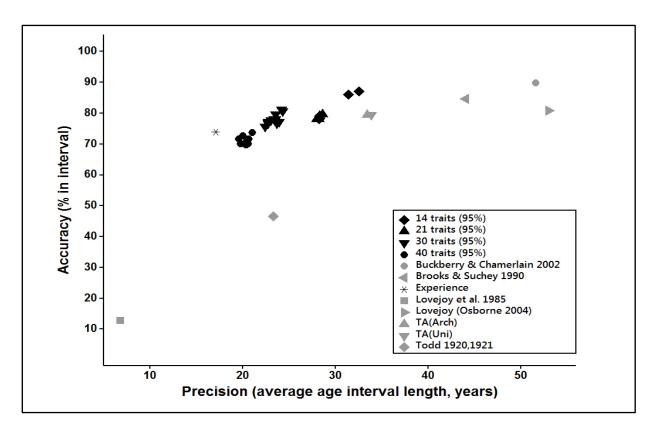


Figure 5.9. Direct comparison of the accuracy and precision of age estimates generated using variations of the new TA procedure (black symbols), commonly used methods for the pubic symphysis and auricular surface, and experience based estimates. Athens Collection, N=201.

The new TA method produces similar levels of accuracy to two of the most commonly used procedures by forensic anthropologists—Suchey-Brooks and Buckberry-Chamberlain—(Garvin & Passalacqua, 2012), but with age intervals that are, on average, half as wide. This is particularly impressive because the new TA procedure does not include any features of the pubic symphysis or the auricular surface. Although these areas include some features that may improve age estimates, particularly in the youngest part of adulthood, they were intentionally excluded from these preliminary investigations. A goal of this dissertation was to test the performance of the transition analysis approach with a larger number of previously unknown or

under-utilized skeletal features. Excluding pelvic joints from the analysis allowed the value of the new features to be unambiguously demonstrated. Modifications of the pubic symphysis and auricular surface features used in existing TA are currently under investigation by the NIJ research team and will be included in the final version of the new TA procedure released to the osteological community. Pubic symphysis and auricular surface trait data consistent with what were collected by the NIJ team were collected for individuals in the Athens and St. Bride's samples and can be included in future tests of new TA.

Bias in Maximum Likelihood Age Estimates

Although informative, plots of overall age estimate accuracy and precision tell only part of the story. In addition to finding a realistic compromise between the width of the intervals and accuracy, an ideal age-estimation method must also produce point estimates that are close to the chronological age of the individual without systematic age-estimation bias. This issue is particularly important for paleodemographic analyzes where overall characteristics of the sample are more critical than those of individual skeletons.

One way to assess the average bias present in a set of age estimates is to fit a linear regression model without an intercept term, often referred to as regression through the origin (RTO). The model is fitted to the estimated maximum likelihood point ages and their associated known ages (Equation 5.0). Because there is no intercept term, the critical feature of the regression is the coefficient for age and its standard error. If the confidence interval for the coefficient contains the value one (1), then the regression line fitted for the sample is statistically indistinguishable from the identity line.

estimated age =
$$coefficient*(documented age) + error$$
 (5.0)

Figure 5.10 shows the RTO coefficients and associated 95% confidence intervals for standard methods (1-7), experience-based estimates (8), and variations of the TA method tested above (9-40). The dotted box indicates the six methods where the fitted regression could potentially have a slope of one. Variations of the new TA method tested with 99% prediction intervals are redundant in this context because changing the prediction interval does not change the maximum likelihood point estimate.

The analysis of slope provides an additional piece to the puzzle, the full picture of method performance is still incomplete. Because fitting a linear regression without the intercept term renders the R-squared value—a common way to assess how well a linear regression model fits the data)—essentially meaningless, an additional plot is needed to assess the scatter of points around the identify line.

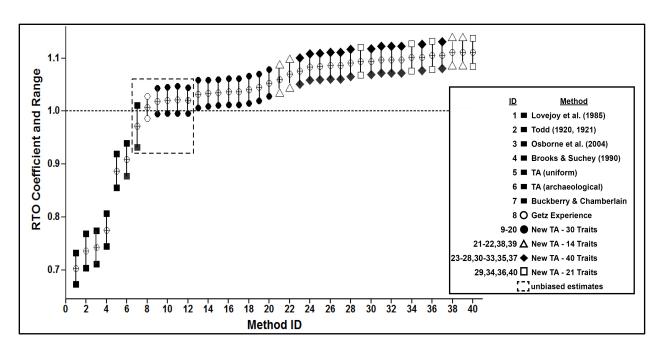


Figure 5.10. RTO coefficients and associated confidence intervals for age estimates produced for the Athens sample (N=201) using standard methods (1-7), observer experience (8), and variations of the new TA procedure (9-40). Confidence intervals that include one (dotted box) indicate that the estimates produced are collectively unbiased.

Figure 5.11 shows the estimated ages produced by three standard methods (circles) with the fitted regression lines (dotted lines) compared to the identify line (solid line). The associated coefficient for each RTO with its associated confidence interval is also shown for each method. Although the fitted regression for the Buckberry and Chamberlain (2002) auricular surface method (c) falls very close to the identify line, the data are a poor fit. Furthermore, as shown in Figure 5.9, although this method demonstrated the highest accuracy of any of the techniques evaluated, it did so by decreasing precision to the point where its age ranges cover most of adulthood. So, while technique offers some improvement over Lovejoy et al. (1985) for the same anatomical area (Figure 5.11, b), its performance is still far from ideal.

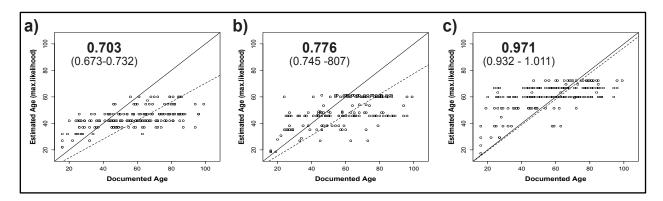


Figure 5.11. Age estimates produced by three commonly used techniques: (a) (Lovejoy, Meindl, Pryzbeck, et al., 1985), (b) (Brooks & Suchey, 1990; Suchey et al., 1988), and (c) (Buckberry & Chamberlain, 2002) with a fitted linear regression (dotted line), regression coefficient, and associated confidence interval. Identify line (solid line) shown for comparison. Athens Collections, N=201.

Similar scatterplots are shown for experience-based estimates, existing TA, and four variations of new TA using different numbers of features in Figure 5.12. All four variations of the new TA procedure collectively overestimate age, with the biggest deviations occurring in old age. This is, however, a long-recognized effect of using a uniform prior distribution (Boldsen et al., 2002). Although this is not a desirable feature of the new procedure, it actually provides support that the new features are collectively providing useful, unbiased age information throughout adulthood. The age estimates generated with the existing Transition Analysis method were also produced using a uniform prior distribution. However, because these ages are based on changes in the pubic symphysis and auricular surface, which fail to keep pace with chronological age after middle age, existing TA still suffers from the problem of underestimating the age of individuals after around age 50 (Milner & Boldsen, 2012c).

Figure 5.13 provides a final look at the results of the standard methods and new TA evaluated in this section and ties together the concepts of accuracy, precision, and age-estimation bias. The known age of each individual is represented by a black dot with the estimated age range and point estimate shown as grey lines. This presentation clearly illustrates the relationship between accuracy and precision and how easily one can be improved at the expense of the other. For example, compare the Todd (1920,1921) and Brooks and Suchey (1988, 1990) methods based on the pubic symphysis and the Lovejoy et al. (1985) and Buckberry and Chamberlain (2002) techniques for the auricular surface.³¹ These graphs clearly show the narrower age ranges and reduction in bias throughout the adult lifespan of the new TA procedure relative to existing TA.

_

³¹ Figure 5.9 demonstrates this relationship in a different form.

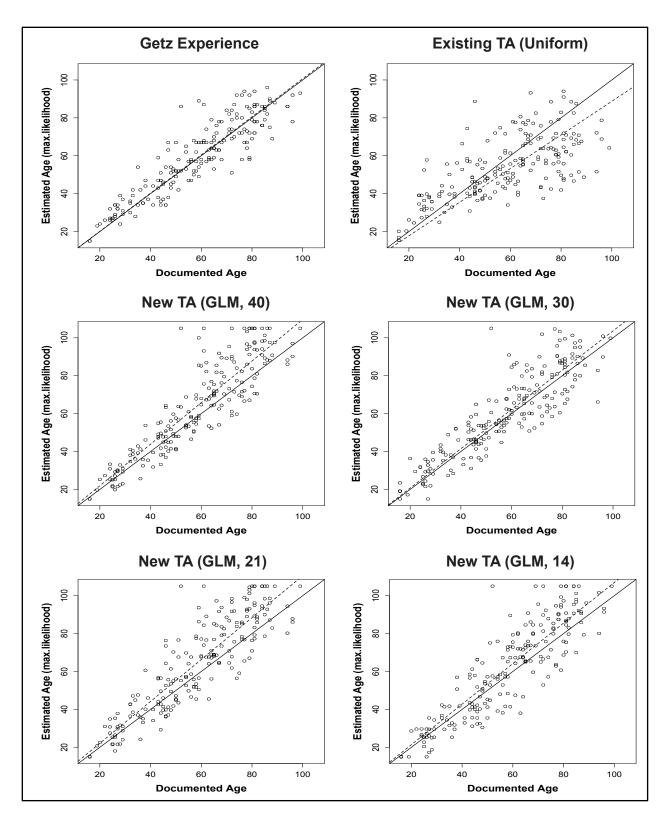


Figure 5.12. Age estimates produced by an experienced observer (Getz), existing TA (Boldsen et al., 2002), and four variations of the new TA procedure using logistic regression probabilities from different numbers of features. The identity line (solid line) and fitted regression through the intercept (dotted line) are shown for comparison. Athens Collection, N=201.

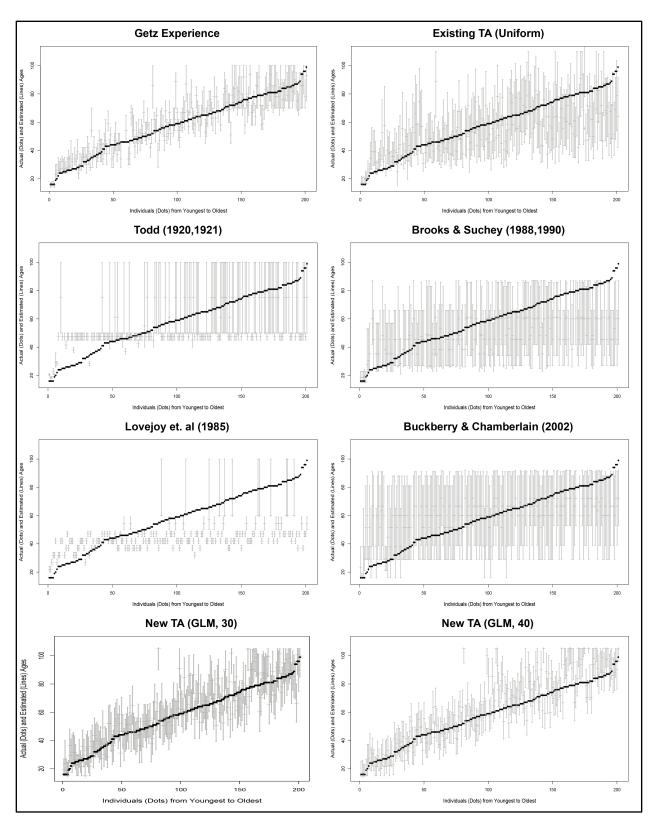


Figure 5.13. Age ranges produced by an experienced observer, existing TA, several standard methods, and the two variations of the new TA procedure. Black dots indicate the known age of each individual, grey lines indicate the age ranges estimated by each method. Athens Collection, N=201.

Evaluation of New TA on a Historical Sample: St. Bride's Crypt

In the previous section, the new TA procedure was applied to modern individuals from the Athens Collection to establish a baseline for the method's performance on well-documented and well-preserved remains. Although these results are promising, adequate performance on modern individuals does not necessarily imply that the method is equally applicable, or will provide the same results, when applied to archaeological samples.

In this section, the new TA procedure is applied to individuals from St. Bride's crypt to assess the performance of the method on a sample that is much different than those in the reference sample. Although the individuals in the St. Bride's sample were interred in a crypt and not directly in the ground, their unique history—decomposition in lead coffins, bombed, exposed to the elements, excavated, shipped for analysis, re-boxed several times, and used frequently for research—has resulted in a sample with a similar state of preservation to what would be encountered in many archaeological contexts. Thus, this sample provides a good gauge for how useful the new traits are for estimating age compared to what can currently be done using existing procedures. Perhaps more importantly, many of the individuals in the St. Bride's crypt sample lived contemporaneously with the individuals who form one of the Danish archaeological samples (Klosterkirke) that is evaluated in Phase 4.

The age estimates generated by standard methods and new TA are compared to those produced in comparable tests with the Athens sample. These data are used again in Chapter 6 to assess the appropriateness of different procedures for modeling population age-at-death distributions. Comparing the age estimates obtained for the St. Bride's sample with those from the Athens Collection indicates whether the method is likely to provide reliable age estimates for archaeological samples.

The effect of skeletal preservation

One factor that may affect the performance of the new TA procedure is the level of skeletal preservation. In many archaeological settings, and often in forensic ones, standard features for assessing adult age, including the pubic symphysis and auricular surface, are missing or damaged. The St. Bride's sample catalog divides the sample into three groups based on preservation of the remains: 1) good, 2) moderate, and 3) poor. Approximately half of the individuals included in the Phase 3 sample were listed as grade 1 (good condition with no cortical bone erosion), while the remaining portion divided between grade 2 (moderate condition with some erosion of bone prominences and articular surfaces), and grade 3 (poor condition with extensive postmortem damage and missing or eroded surfaces throughout).

With this in mind, the percentage of the Phase 3 St. Bride's sample that could be evaluated using the pubic symphysis (light grey), auricular surface (dark grey), observer experience, existing TA, and the new TA procedure (black) are shown in Figure 5.14. As is often reported in other studies, the auricular surface is more often preserved than the pubic symphysis. For this sample, in cases where only part of the scoring area was present or portions were slightly damaged, multiple scores were assigned (e.g., Suchey-Brooks Phases 5-6) to account for the possible features that may have been present in the damaged area. Therefore, the percentages of individuals who could be scored using each method are slightly higher than they would be if only perfectly preserved individuals were scored. Although assigning multiple phases allows something to be said about partial or damaged remains, it often widens the age range to the point where it covers almost all of the adult lifespan, which provides little useful information.

The high percentage of individuals that could be scored with existing TA, despite the high number of damaged pelvis joints, results from the use of separately scored components. Existing TA has a maximum of 32 scores (9 for each auricular surface, 5 for each pubic symphysis, and 5 cranial suture scores). New TA has a maximum of 40 scores, which does not include any features from the pelvis joints, and does not include bilateral scores at this point In the data presented here, a minimum of five scores had to be present for both existing TA and new TA to be counted in the percentage of individuals who could be evaluated. Just over 95% of the individuals in the sample (160/168) could be scored using the new TA procedure, while only 46% (77/168) could be assessed using the method most commonly used by forensic anthropologists and archaeologists to evaluate the pubic symphysis (Brooks & Suchey, 1990; Falys & Lewis, 2011; Garvin & Passalacqua, 2012; Suchey et al., 1988).

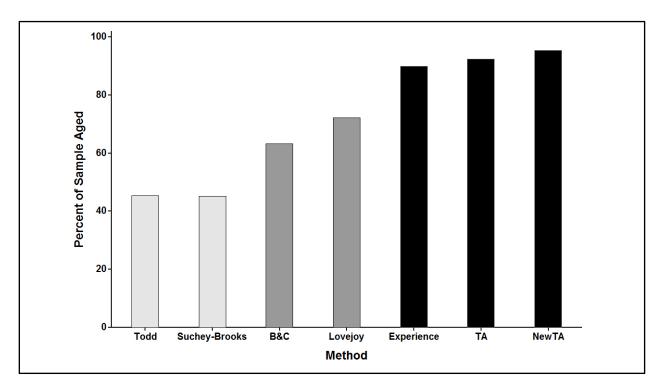


Figure 5.14. Percentage of the St. Brides crypt sample (N=168) that could be aged using methods for the pubic symphysis (light grey), auricular surface (dark grey), and those that use features from multiple skeletal elements (experience-based estimates, existing TA, and new TA) (black).

It is usually assumed that any information about a skeleton is better than no information at all, which is why multiple stages are often assigned to individuals, even though the resulting ranges span most of adulthood. Because many skeletons in archaeological samples are likely to have at least some features that cannot be scored, the number of features required to produce reasonable age estimates, individually or in aggregate, becomes a key issue. Figure 5.15 shows the distribution of the number of features that could be scored on each skeleton in the St. Bride's sample. Only a single individual had all 40 features intact. The majority of individuals in the sample have between 20 and 40 observable features, which previous tests with the Athens sample indicate may be enough to produce reasonable estimates.

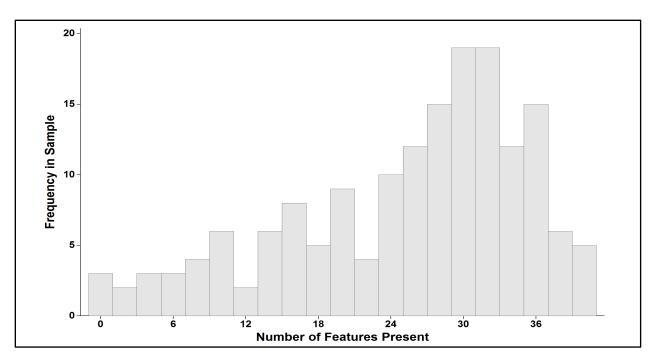


Figure 5.15. Distribution of the number of scorable features of each individual in the St. Brides sample (N=168).

Tests with the Athens sample indicated that using a reduced set of 30 features produces the best compromise between accuracy and precision. However, as shown in Figure 5.15, if only individuals with 30 or more features were used, a large fraction of the St. Bride's sample could not be evaluated. Additionally, it is unclear whether different groups of 30 features would provide comparable accuracy and precision to the specific group of 30 features that were originally tested. Fortunately, Figure 5.16 (and the scatterplots previously presented in Figure 5.12) demonstrate that estimates produced using sets of 14, 21, 30, and 40 features are similar. Although, as would be expected, the highest level of variation is seen in estimates generated using the smallest set of features, the estimates do not differ substantially among the groups.

The age estimates produced for the St. Bride's sample provide additional evidence to suggest that threshold of approximately 20 features may be appropriate for evaluating archaeological samples. Figure 5.17 shows the lengths of the age intervals generated using different number of features. The pattern using probabilities based on a combined reference sample (a) or either of the death cohorts (b) is the same—the length of the interval narrows as additional features are added until around 25 features, when no additional improvement occurs. Very narrow range lengths most often occur at the extremes of the age distribution where all traits that could be observed are scored as either present or absent. Other exceptions to the general trend potentially result from the exaggerated effect that correlated features may have when only a few traits are available for analysis. For example, an individual who could be scored for fifteen features, ten of which are in the spine, may have an abnormally narrow age interval resulting from the use of more highly correlated traits. Additional testing using simulated groups of features is needed to determine what other factors may be influencing age-interval length and if the pattern of decreased interval length with an increasing number of traits, to a point, occurs consistently in other samples.

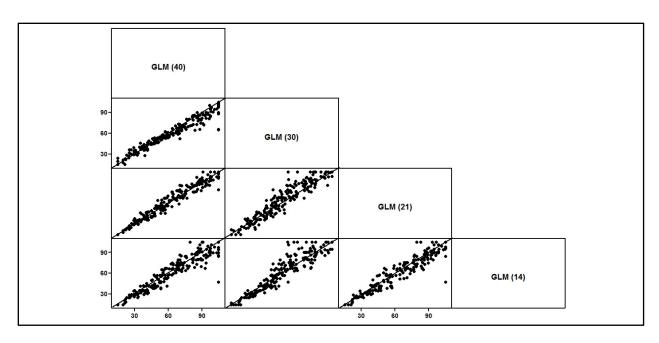


Figure 5.16. Comparison of age estimates produced for the Athens sample (N=201) using combined reference sample logistic probabilities from different numbers of features.

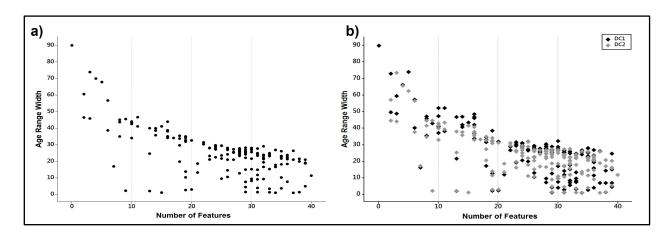


Figure 5.17. Relationship between the number of features used to estimate age and the length of the interval produced using (a) combined or (b) time period-specific trait probabilities and a 95% prediction interval. St. Bride's Crypt sample (N=168)

Accuracy and precision

Similar to what was done for the Athens Collection, the accuracy and precision of several variations of the new TA procedure are first compared with the performance of standard methods. Figure 5.18 compares six variations of the new TA procedure (solid symbols) with other commonly used methods (open symbols), and experience-based estimates (asterisk). The overall pattern is the same as was seen with modern individuals. Existing methods fall somewhere along the continuum from high precision with low accuracy (lower left) to high accuracy with low precision (upper right), with existing TA, new TA, and experience-based estimates in the middle and slightly offset from this pattern.

Minor variations in the pattern of accuracy and precision for standard methods between the two collections primarily result from differences in the age distributions of each sample. For example, the accuracy of Todd (1920,1921) (open diamond) was shifted from approximately 25 years in the Athens sample to 45 years in this application with a corresponding increase in accuracy. This is because the individuals in the St. Bride's sample who were scored for this feature have an older age distribution than those that could be scored in the Athens collection. Therefore, a greater number of individuals fell into the terminal 50+ years age category, which resulted in a higher average age-interval length. Another factor contributing to minor differences between the results of the two samples is the smaller number of individuals that could be scored using some of the techniques because of postmortem damage (see Figure 5.14).

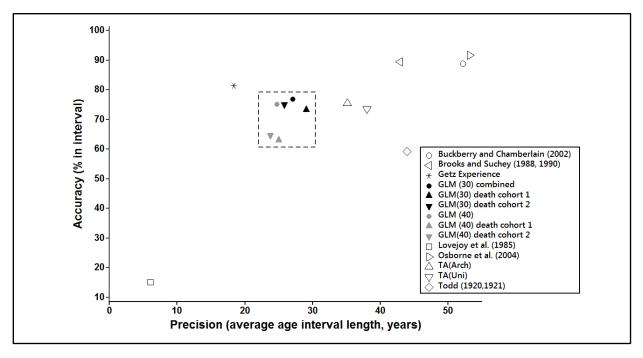


Figure 5.18. Accuracy and precision of age estimates generated using new TA procedure variations (dotted box), commonly used methods for the pubic symphysis and auricular surface (open symbols), and experience-based estimates (asterisk) for the St. Bride's Crypt sample (N=168).

The variations of the new TA procedure (dotted box) perform exactly as anticipated based on the previous tests using modern individuals. The accuracy and precision produced for the St. Bride's sample using logistic probabilities for 30 and 40 traits is highly similar to the same method variations in the Athens sample. This can been seen by comparing the results presented in Figure 5.9 with those in Figure 5.18. As in the Athens sample, using 30 features produces slightly higher accuracy that using all 40 possible traits, but this effect is not as marked as was with the modern, well-preserved individuals. This is likely because many individuals in the St. Bride's sample are incomplete. Allowing for a maximum of forty features increases the average number of features per individual that are used while rarely actually including that number (see Figure 5.15).

For both the Athens and St. Bride's samples, using reference data from the entire Phase 1 sample results in slightly better age estimates overall than those produced using time period specific trait probabilities. However, there is a potentially interesting deviation in the results obtained when applying probabilities from both death cohorts to each sample. In the Athens sample of modern individuals, using probabilities from death cohort 2 (post-1981) slightly reduces the overall accuracy, but produces very similar results to using the combined reference sample. Using death cohort 1 (pre-1953), however, resulted in the highly unusual combination of wider age intervals with a decrease in accuracy. In the St. Bride's sample, the use of either death cohort produces slightly lower accuracy than using the combined sample, but neither death cohort shows the dramatic decrease in accuracy seen in the Athens sample.

This is potentially the result of at least two inter-related factors. First, the size of each of the death cohorts is necessarily smaller than that of the combined sample. These sample size differences produce slight variations in the probabilities produced for each trait (Appendix B), which will have a different overall effect on the performance of the method on a sample depending on the underlying age structure of that sample. Second, many of the skeletons in St. Bride's sample are incomplete and, typically, the use of fewer traits results in a wider age interval with correspondingly higher accuracy. This could mean that regardless of the reference sample used—death cohort 1 or death cohort 2—a significant number of the log-likelihood functions for the individuals in the St. Bride's sample are so wide that they cannot help but be accurate because they are based only on a few features.

The high frequency of remains that are at least partially damaged in the St. Bride's sample also creates variability in the age intervals produced for standard methods, including existing TA, and the new TA procedure. In Figure 5.19, each individual in the sample is represented by a black dot with the estimated age range and point estimate as grey lines. As discussed for the Athens sample, this presentation illustrates the typical relationship between precision and accuracy. In this case, the variability introduced by partial remains can also be seen by the wide age ranges that appear erratically throughout adulthood, particularly for existing TA (top right) and new TA (bottom row). If this figure is compared to the similar figure produced for the much more complete individuals of the Athens Collection (Figure 5.13) the effect of sample preservation on the age ranges produced using each method becomes even more clear.

Point estimates and age-estimation bias

Figure 5.18 demonstrates that the overall accuracy and precision of new TA in the St. Bride's sample is similar to that obtained using a well-preserved modern sample, while Figure 5.19 shows that the overall patterns for the entire adult lifespan are similar. Both issues are critical to understanding how each method can be expected to perform when applied to different sample. For forensic applications, the point age-at-death estimate is less important than the age range and its associated level of confidence. However, for archaeological applications, the performance of the point estimates, including their error and bias across the lifespan, is of particular importance. Although it is not a statistically or theoretically advanced technique, these estimates are, by far, the most frequently used way to generate population age-at-death distributions from archaeological data (Hoppa & Saunders, 1998; Jackes, 2000).

In Figure 5.20, the point estimates of age (means, midpoints, or maximum likelihoods) produced by the best-performing standard methods (Brooks and Suchey 1990, Buckberry and Chamberlain 2002), existing TA, observer experience, and new TA are compared to the known ages-at-death for the same individuals. The identity line (solid line) and fitted linear regression through the origin (dotted line) are provided to facilitate comparisons between the methods. The

overall patterns in the estimates obtained for the St. Bride's sample for each method are essentially identical to those obtained for the Athens sample using the same techniques. The point estimates produced by standard methods (top row) show a general increase with age, but are a poor fit for either the identity line or the fitted regression. Existing TA (middle row) allows age to be estimated for more of the adult lifespan, but produces a large scatter of points when used with incomplete remains and collectively underestimates age. The new TA procedure (bottom row) performs as expected based on previous tests, with one discrepancy. A greater number of individuals in the St. Bride's sample were estimated as having ages at the extremes of the distribution—15 and 105 years—than in tests with the Athens sample. These individuals appear as a line of dots at the top and bottom of the new TA graphs in Figure 5.20.

Two explanations are probably needed—one for the number of individuals estimated as 15, and the other for the people at age 105. The use of partial remains likely has a strong influence on both. First, as discussed for the Athens sample, the group of individuals aged near the maximum limit is partly an artifact of the use of a uniform prior distribution where it is assumed that a skeleton is equally likely to be any age in the adult lifespan. Any individual with many "oldage traits"—features with transitions in the eighties, nineties, and above—will be estimated to be in the far upper extent of the lifespan, even though it would be far more likely in any population to find an 80-year-old with those scores than a 100-year-old individual. This problem is exacerbated in incomplete individuals because it becomes more likely that all or many of the traits in older individuals will be scored as "present" when there are fewer features for analysis.

Second, two plausible explanations exist for the overestimation of individuals as 15 years of age—one theoretical and one practical. It is possible individuals under the age of 30 years in the St. Bride's sample may be slightly delayed in their development relative to the individuals in the Phase 1 reference sample. This would result in a greater number of individuals having all young trait scores and being placed in the youngest age category. However, if this explanation were the most likely, it is reasonable to assume that the ages produced using probabilities from the two different death cohorts would result in different numbers of individuals estimated to be in the youngest age category, which did not occur. The use of either death cohort, or the combined sample, produces the same anomalies at the youngest and oldest ages. Although this is not definitive evidence that differences in the timing of age-related changes do not exist between the samples, particularly considering the low number of young individuals in death cohort 1, it indicates that other factors may be preliminary responsible in this case.

Because this effect at the youngest ages is not seen in the Athens sample and does not change with the reference sample used, it is likely that the fragmentary nature of the remains and the particular mix of traits available in each individual are primarily responsible. As for older individuals, when fewer traits are available for analysis it becomes more likely that all of the binary features will fall into a single category, which results the youngest possible estimated age. This effect is greater for young individuals because there are currently very few features with transitions at the youngest ages in the new TA procedure. This is because the focus of this work was to identify features to improve estimates in middle and old age and to demonstrate the utility of new skeletal features independent of information from traditional indicators. This means that the young-age features of the pelvic joints that work well in traditional methods for individuals under the age of 40 were not included in these estimates. The NIJ team has made changes to the pubic symphysis and auricular surface features from existing TA, and they will be included in the final version of the new TA method. The inclusion of these features, in addition to new NIJ-defined traits in other parts of the skeleton with transitions in young adulthood, should resolve the age-estimation issue for young adults.

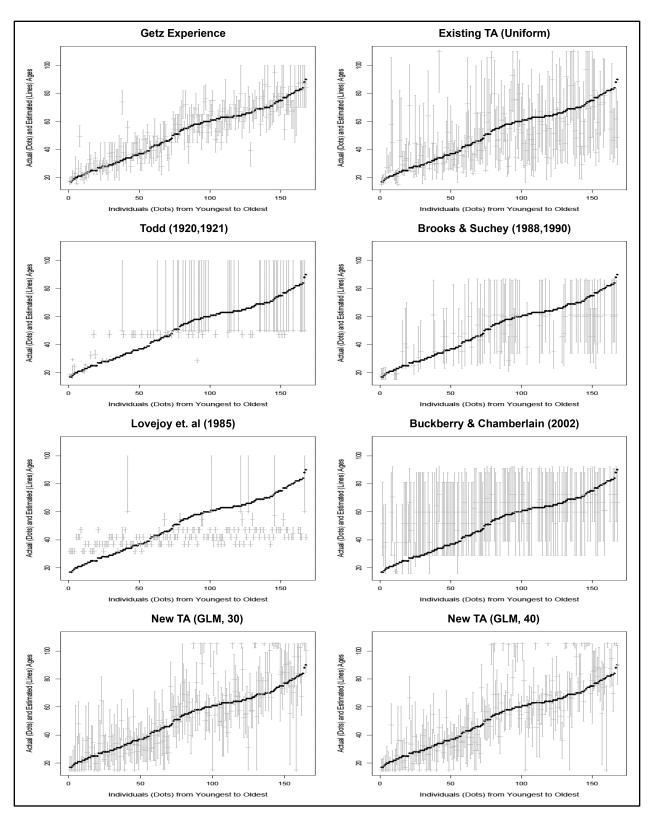


Figure 5.19. Comparison of the age ranges produced by an experienced observer, existing TA, several standard methods, and two variations of the new TA procedure for the St. Bride's Crypt sample (N=168). Black dots indicate the known age of each individual and grey lines indicate the range estimated by each method.

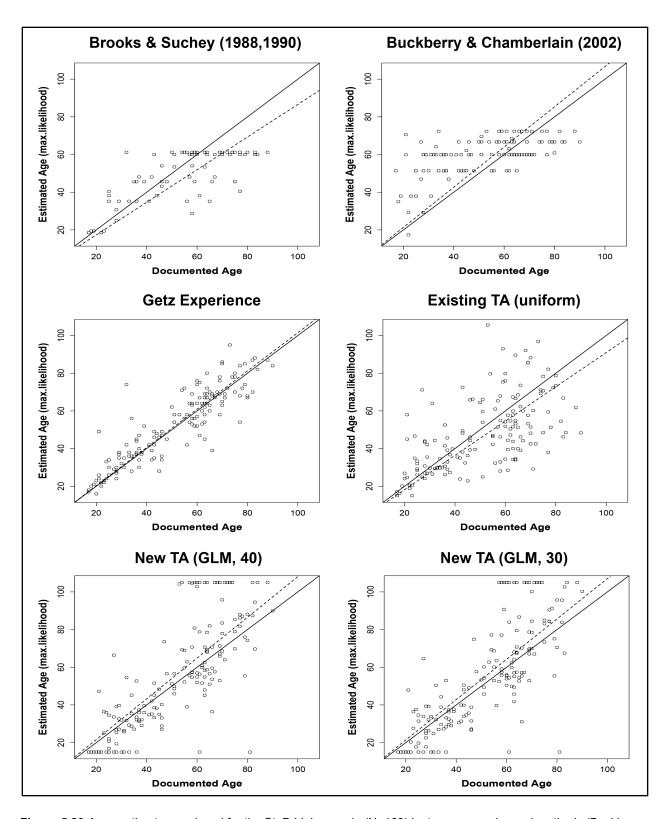


Figure 5.20 Ages estimates produced for the St. Bride's sample (N=168) by two commonly used methods (Buckberry & Chamberlain, 2002) an experienced observer, existing TA (middle row), and two variations of the new TA procedure using logistic regression probabilities for different numbers of features (bottom row). The identity line (solid line) and regression fitted through the origin (dotted line) are shown for comparison.

Point estimates and age-at-death distributions

The final element needed to evaluate the potential performance of standard methods and new TA for archaeological applications is the age-at-death distributions that result from grouping point age estimates into age categories. Each panel of Figure 5.21 shows the documented age-at-death distribution (solid line) for individuals in the St. Bride's sample overlaid with distribution of estimated ages (dashed line) produced using each method.

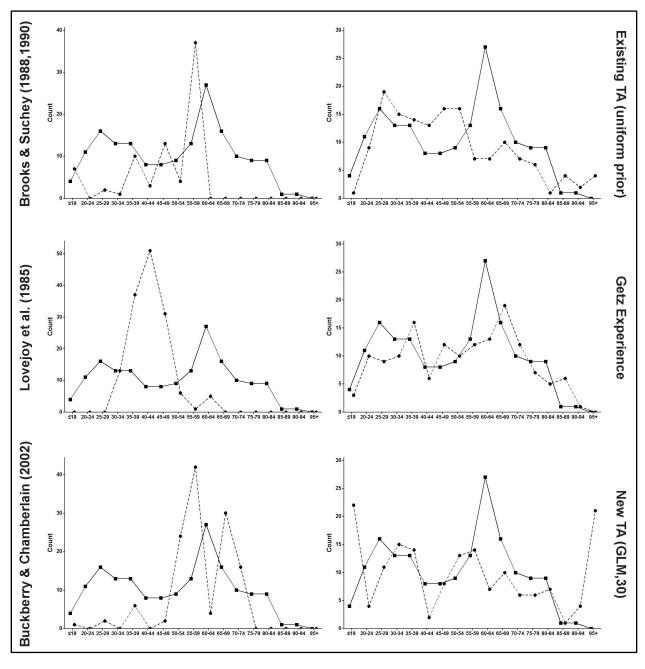


Figure 5.21 Known (solid line) and estimated (dashed line) age-at-death distributions for the St. Bride's crypt sample produced using standard techniques (left column), existing TA, observer experience, and new TA.

Based on these results, it is not difficult to see how using the (a) Suchey-Brooks pubic symphysis or (b) Lovejoy et al. auricular surface methods would result in an archaeological age-at-death distribution with few individuals living beyond around 60 years of age (e.g., Bullock et al., 2013; Nagaoka & Hirata, 2008). The Buckberry-Chamberlain technique (c), also based on the auricular surface, produces a slightly different pattern with too few individuals under the age of 40 and too many between 50 and 80. This is potentially why a number of studies have found this method to be most appropriate for samples of older individuals (e.g., Hens & Belcastro, 2012; Nagaoka & Hirata, 2008; San Millán, Rissech, & Turbón, 2013). These findings also support the long-discussed idea that the oddities observed in many archaeological age-at-death distributions, particularly the large number of people dying in middle age, are the result of the peculiarities of the methods used and do not reflect the true features of the mortality sample (Bocquet-Appel & Masset, 1982; Howell, 1982; Konigsberg & Frankenberg, 1992).

Regardless of the reason, it is clear that the age distributions produced using the traditional procedures evaluated here do not adequately represent the true age distribution of the St. Bride's sample (Figure 5.21). The age-at-death distribution produced using estimates from existing TA, however, shows that there is hope. By scoring features of the pubic symphysis and auricular surface as individual components and combining them in a different statistical framework based on a reasonably large sample, the correspondence between the estimated and known age distribution improves dramatically; however, despite this improvement, existing TA is not free from age-estimation bias. The slight over- and under-estimation of age before and after around age 50 respectively can be seen in the estimated age distribution (dashed line) relative to the true distribution (solid line) in the fourth row of Figure 5.21. As previously discussed, this is primarily the result of the method's reliance on cranial and pelvic features where morphological change fails to keep pace with chronological age after around age 50 (Milner & Boldsen, 2012c).

The new TA method takes the statistical framework of existing TA, excluding the correction for correlated features, and improves its performance with a suite of new traits. Although the new TA method overestimates the number of individuals in the very youngest and oldest age categories, if these same spikes are seen in archaeological samples, it can be assumed that the same processes are at work. This assumption can be verified in the youngest ages by crosschecking the ages produced for the youngest category with the overall pattern of skeletal development for these individuals. The new TA procedure does not include the majority of sites of epiphyseal fusion because most provide information only for children and teenagers, and would therefore contribute little to an adult age-estimation method. These overall pattern of development and age-at-fusion for these sites is, however, relatively well documented and can be used to assess the number of individuals in the archaeological sample that likely fall between 15 and 20 years of age.

In the oldest ages, observer experience is used to assess the plausibility of the age estimates. Although experience is not a scientifically valid technique because the uncertainly of the estimates cannot be quantified for any particular case, in these investigations of the new TA procedure it has been shown to be a reliable tool for estimating age in multiple samples. Because the age-at-death distribution produced using experience most closely matches that of the known distribution, it is used to help understand the performance of the new TA technique and assess possible sources of error, such as missing data, taphonomic modifications of bone texture, disease processes, and individuals where the documentation may not match the remains. Ultimately, it is hoped that the addition of new features to this procedure will result in a method that reproduces or exceeds what can be done by a highly trained observer.

Phase 4: An Archaeological Application of the New TA Procedure

Taken together, these results provide strong support for the efficacy of new TA in archaeological contexts. Applying the new method to the St. Bride's sample allowed between thirty and fifty percent more of the sample to be aged compared to using standard methods for the pubic symphysis or auricular surface alone. Approximately the same number of individuals were assessed as with the existing TA method, which is based only on cranial sutures and pelvic joints, but with greater accuracy, higher precision, and less bias. The use of probabilities from time-specific reference samples had no significant affect on the accuracy or precision of the method and a combined reference sample produced the best results. Although using a uniform prior distribution resulted in an unrealistic number of individuals in the oldest age categories, the overall age-at-death distribution generated using five-year age categories was a good approximation of the true sample distribution.

In Phase 4, the new TA procedure is applied to two contextually well-documented Danish archaeological samples—Ole Worms Gade (1100–1500 CE) and Klosterkirke (1500–1800 CE). Individuals from these samples are evaluated using standard age-estimation methods, existing TA, and new TA. Data from the St. Bride's sample are used to evaluate the appropriateness of these methods for generating morality distributions from archaeological samples. This application demonstrates the degree to which the use of different age estimation techniques and methods of modeling age distributions potentially impacts the conclusions drawn from archaeological data and the extent to which existing samples should be re-evaluated based on these findings.

CHAPTER 6: PHASE 4—ARCHAEOLOGICAL APPLICATION: THE DANISH CITY OF HORSENS IN THE LATE MEDIEVAL & EARLY MODERN PERIOD

Skeletons, archaeological artifacts, and written records provide three overlapping sources of evidence for the reconstruction of population histories (Cannon, 1995; W. Petersen et al., 1975). Because no single source is without gaps or biases, multiples lines of evidence are needed to identify and mitigate individual shortcomings (Cannon, 1995; D. Ubelaker, 1995). In the late medieval period in Europe, the number of written and pictorial sources describing the functions of towns dramatically increased (H. Anderson, 2007), including law codes and privileges, records of meetings of town governments, tax registers, customs accounts, deeds, and assorted other documents (Dahlbäck, 2008). These sources can complement archaeological data in investigations of social and demographic patterns (H. Anderson, 2007).

During the Middle Ages, the number of towns grew rapidly, although they remained small, and, by the fourteenth century, Denmark was the most densely urbanized of the Scandinavian countries (Benedictow, 2008; Dahlbäck, 2008). This growth was soon halted, however, by a period of unusual cold that resulted in food shortages, which were soon exacerbated by outbreaks of plague (H. Anderson, 2007). No towns were founded in Denmark in the first half of the fourteenth century (H. Anderson, 2007). After the massive depopulation of the Black Death starting around 1350 CE and lasting to ca. 1500, a general pattern of accelerated urban growth emerged that has continued to today (Johansen, 2002).

The expansion of cities could be related to an influx of individuals into urban areas, particularly young-adult males and couples, increases in life expectancy, or a combination of both. Although few written sources can shed light on internal migrations during this transition period (Johansen, 2002), a combination of plague epidemics, cooler climates, changes in land ownership and management practices and political unrest undoubtedly contributed to these migration patterns that shaped the country as it is today (H. Anderson, 2007; Gissel, 1981; Vahtola, 2003). Increased mobility, dietary changes, and reduced disease load, along with other factors, substantially contributed to the shifting demographic patterns between the late Middle Ages and Early Modern period (Orrman, 2016; Sogner, 2016).

Medieval demography can generally be characterized as having high birth rates and high death rates, with as many as half the children dying before age seven and very few individuals living over the age of 60 (Vahtola, 2003). Historical records indicate that mean life expectancy at birth in the Middle Ages was between 18-23 years at birth, or 15-18 years from age 20, and had increased to between 30 and 40 years at birth by the Early Modern period, with females dying slightly earlier in both cases (Johansen, 2002; Vahtola, 2003). Similarly, Benedictow (2008) gives a mean life expectancy at birth between 20 and 25 years in the Middle Ages that increased to around 35 years by the mid-eighteenth century. This implies substantial changes to the demographic structure of the population between medieval to early modern society (Benedictow, 2008). It is debated, however, whether increases in fertility or declines in adult mortality played the biggest role in the rapid growth (Benedictow, 2008; Livi-Bacci, 2012) Skeletons are an important source for evaluating reconstructions based on documentary sources (Walker, 1995), and may be able to help shed light on complex demographic changes.

Unfortunately, the extensive literature investigating standard age-estimation techniques (see Appendix A) indicates that the application of these methods to archaeological samples would almost certainly yield biased results. This is further supported by the results presented in

Chapter 5 of this dissertation. Using existing techniques, it is likely that changes in life expectancy could not be detected because age ranges for individual skeletons are large, often thirty to fifty years or more, and all individuals over 50 are lumped into a single category. Any real demographic shifts would be further masked by the errors introduced by assigning individuals to age categories using methods that have been repeatedly shown to have low accuracy and systematic age-estimation bias.

The focus of this phase is to evaluate the performance of commonly used age-estimation techniques, including existing TA, compared to the results produced by the new TA procedure. Information from the exploratory tests conducted with known-age samples in Phase 3 will be used to interpret the results of these applications. The more precise and collectively unbiased age estimates produced by the new TA method should allow differences to be detected in the age-at-death distributions of the two samples, if they indeed exist. Contextual information about the two sites and their excavations is used to identify potential sources of bias in the sample to aid in the interpretation of differences in the mortality profiles, which may cause them to differ from historical sources, even with adequate age estimates (Saunders et al., 1992).

Phase 4 Samples

In this phase, two samples were selected to represent the Middle Ages and Early Modern period in Denmark where we would expect to see changes in mortality profiles based on historical documentation—Ole Worms Gade (Ole Worms Street) and Klosterkirke (Monastery Church). Importantly, these sites were located less than 1 km apart (Figure 6.0), so an assumption can be made about the genetic and cultural similarity of the populations that contributed to the two samples. Both samples are also sufficiently large enough to provide adequate samples of individuals that can be assessed using standard methods, existing TA, and new TA, so that reasonable mortality distributions can be produced.

Table 6.0 summarizes key information about the Ole Worms Gade and Klosterkirke samples used in this phase. All individuals from each site, with the exception of several on long-term loan to other institutions at the time of data collection, were evaluated.³² Any individual that was estimated to be 16 years or older based on dental development and the overall pattern of epiphyseal fusion in the skeleton was included in the sample.³³ In a small number of cases, sets of remains were excluded from the sample because they appeared to contain elements from two or more individuals. In cases where commingling was suspected, the original excavation and skeletal analysis records were consulted before making the decision to exclude the individual. Bones that were not found in the context of a primary burial (i.e., loose finds) were not analyzed because of the inaccuracy in age estimates based on only a handful of features and the potential for duplicating individuals in the final sample.

-

³² Seven individuals were on loan and therefore unavailable for analysis. According to the osteology report for the collection, two were children under the age of 13, and five were adults spaced out between 24 and 57 years of age based on experience-based estimates.

³³ Juveniles will not be considered in this application. However, individuals under the age of 16 were assessed to verify the age assigned during the original osteological analysis of the sample by ADBOU staff. These data are available for future work.

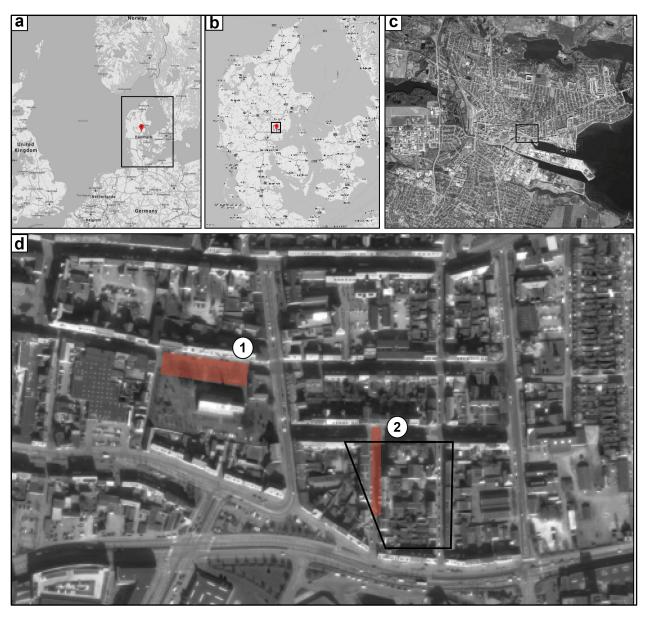


Figure 6.0. Location of Phase 4 archaeological samples from Horsens, Denmark at different scales: (a) Northern Europe and Scandinavia, (b) Denmark, (c) city of Horsens, and (d) approximate excavation locations. The rectangular box on each map (a-c) indicate the area enlarged in the next frame. In frame (d), the highlighted areas are the two sites—(1) Klosterkirke and (2) Ole Worms Gade. The trapezoidal box near (2) indicates the likely extent of non-excavated portion of the cemetery. Maps modified from (Google, 2016).

Table 6.0. Archaeological collections used in Phase 4

Collection	Years of Use	Majority of Burials	N	М	F	Unknown
Ole Worms Gade (HOM 1649)	1100 - 1536	1350 - 1536	317	158	129	30
Klosterkirke (HOM 1272)	1536 - 1856	1750 - 1800	166	85	78	3

HOM 1649—Ole Worms Gade (Ole Worms Street)

Archaeological excavations at Ole Worms Gade took place between 2007 and 2009 after skeletal remains were discovered during construction related to water and sewer lines (Klemensen, 2009). Only the portion of the cemetery under the modern road was excavated; however, based on reports of other archaeological remains found in the area, the cemetery likely extends slightly westward along a road shown in maps from the Middle Ages, and east at least one modern block under apartments and other buildings. The full extent of the cemetery area is estimated to be around 9,400m², of which only approximately 580m² were excavated (Klemensen, 2009). Figure 6.0, panel d(2), shows the location of the excavated area (red shaded area) and the likely extent of the cemetery (black outline).

The Old Worms Street cemetery was in use since the twelfth century, but essentially all the burials date between 1350 and 1536 CE. In total, there are 401 individuals from primary burial contexts—278 adults (16 years or older) and 132 children (Pedersen & Boldsen, 2010). Only the area under the road could be excavated, so more than 55% of the individuals in the sample are represented by a third or less of a skeleton³⁴ (Pedersen & Boldsen, 2010). In addition to constraints imposed by street and surrounding buildings, at least a third of area under the street had been disturbed over the previous century by the installation of sewer, water, and gas lines, as well as at least two sets of heating pipes (Klemensen, 2009). Although more than 75% of the sample is listed as in good or excellent condition, many sets of remains contain only a few well-preserved bones.

HOM 1272—Horsens Klosterkirke (Monastery Church)

Excavations at the Klosterkirke site took place between November 2006 and March 2008 (Tarp & Boldsen, 2010). The site is named because of its close proximity to a Franciscan monastery that was likely demolished in the late 1500s (C. G. Petersen, 2007). Burials at the site began in 1536, but by 1825 the area had become so crowded that it could no longer handle the deaths of a growing city. In 1835 a new cemetery was founded in another location and the Klosterkirke cemetery was abandoned several decades later (C. G. Petersen, 2007). Although the site was officially in use between 1536 and 1856, the majority of excavated burials are believed to be from the second half of the eighteenth century (Tarp & Boldsen, 2010). The site contained 221 primary burials, including 178 individuals who were 16 years of age or older³⁵ (Tarp, 2010). Approximately three-quarters of the sample is listed as being either moderately or well preserved with, on average, about half of each skeleton present (Tarp & Boldsen, 2010). Of these, 166 individuals were available at the time of data collection and were suitable for analysis.³⁶

_

³⁴ The remaining portions of which remain *in situ* under the edges of the excavation area or in inaccessible areas of the site, such as under existing pines

the site, such as under existing pipes.

The age of individuals in primary burials was assessed using epiphyseal fusion, dental development, the cranial sutures, pubic symphysis, and auricular surface, as well as additional features of the femur. Each individual was assigned an age range. If the midpoint of this range was 16 years or older, the skeleton was considered to be an adult (Tarp & Boldsen, 2010).

Bight individuals were on loan to another institution at the time of data collection and were unavailable for analysis.

³⁰ Eight individuals were on loan to another institution at the time of data collection and were unavailable for analysis Several skeletons were not included in the sample because of possible commingling of remains.

Methodological Considerations

Traits to include

Tests with the Athens sample in Phase 3 indicate that using a reduced sample of 30 features, rather than all 40 traits, improved the compromise between accuracy and precision and slightly reduced the collective bias of the estimates. Using a smaller group of 21 or 14 traits also produced essentially unbiased results, but increased the variability in the point estimates.³⁷ In well-documented samples, the issue of how many features to include in an age estimate is, at this point, simply an interesting exercise. For archaeological investigations, however, this issue has potentially serious implications for the use of the method because skeletons are often incomplete. Artificially restricting the number of features used may result in a substantial loss of data and introduce additional biases into the sample.

In both of the Phase 4 samples, more than 50% of the individuals are represented by less than a third of the skeleton, while another 25% have less than two-thirds of the skeleton (Pedersen & Boldsen, 2010; Tarp & Boldsen, 2010). Figure 6.1 shows the number of new TA features used to estimate age for the individuals in each archaeological sample when either all 40 or a subgroup of 30 traits were allowed as possible scores. In these samples, using the 40 features results in a greater number of individuals with more traits that can contribute to age estimates (Figure 6.1, dark grey bars with lines). In this situation, the benefits of using a large suite of features, outweigh the possible drawbacks of including too many. Particularly because only a single individual out of all 485 could be scored for all 40 traits and over 90% of each sample could be scored for fewer than 30 of the 40 possible features.

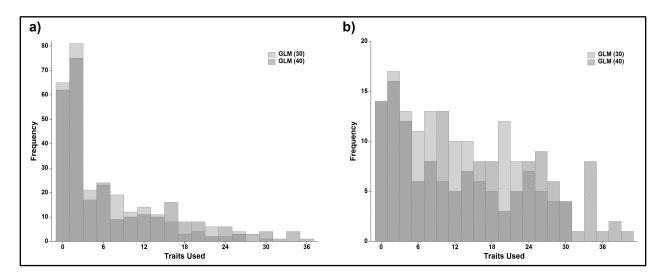


Figure 6.1. Distribution of the number of features available for estimating age for the individuals in each sample—(a) HOM 1649: Ole Worms Gade (N=317) and (b) HOM 1272: Klosterkirke (N=168)—when either 30 or 40 traits were possible. Note the difference in the y-scale between the two charts.

 $^{^{}m 37}$ These results were summarized in Figures 5.9 and 5.12 in the previous chapter.

The differences in the distribution of the number of traits present between the samples result from the distribution of the new TA features in the skeleton combined with differences in the sites themselves. The Ole Worms Gade sample (Figure 6.1, a) has a much larger number of individuals with fewer than five features present. Unfortunately, despite significant effort to select features to represent the entire skeleton, there are currently only a small number of new TA features from the knee down. Because of the orientation of individual of these individuals under the road surface, a substantive portion of the sample is represented only by lower limbs or isolated feet. In contrast, in the Klosterkirke sample (Figure 6.1, b), the distribution of features present represents the more random loss of bones that results from the long-term use of a small area for many burials. Although graves could be excavated in their entirety, over many years newer burials were laid on top of and cut into older ones, which resulted in a large number of loose and commingled remains in the grave fill that could not be scored.

The effect of too few traits

In Phase 3, it was demonstrated that Getz's experience-based estimates have a strong and unbiased correlation with documented age in both the St. Bride's and Athens samples. Although, at this time³⁸, expert age assessments estimates cannot be used in a formalized way to evaluate samples, they are used in this section to assess whether the new TA procedure is functioning in a similar way in both the known-age and archaeological samples. Of the two archaeological samples evaluated in this phase, Klosterkirke is likely the most similar to the known-age individuals from St. Bride's crypt assessed in the previous chapter. Although not an exact match, the two samples contain individuals of similar ancestry who lived in an urban environment in northern Europe during the same time period. Therefore, comparisons of these two samples—one with documented ages and one where they are unknown—are first used to test the efficacy of the new procedure.

Figure 6.2 compares experience-based estimates to the maximum likelihood estimates produced by new TA using the logistic trait probabilities of 40 features from the full Phase 1 reference sample (i.e, one logistic model fitted per trait to a combined sample of all males and females from both death cohorts) for the St. Brides and Klosterkirke samples. The two panels for each sample show the same age estimates but individuals who were aged using fewer than five or fifteen features are shown in red. In the St. Bride's sample (Figure 6.2, top row), only a single individual had fewer than five features present. Removing individuals with fewer than fifteen features would eliminate several outliers, but would also eliminate some individuals where age could be reasonably estimated. Excluding individuals based on this cutoff would not substantially improve the performance of the method at the sample level.

Age estimates for the archaeological Klosterkirke sample (HOM 1272) (Figure 6.2, bottom row) show similar patterns, although more individuals fall into both the under-five and under-fifteen trait thresholds because many of the skeletons are incomplete. Removing individuals in both categories would mitigate problems seen at both the youngest and oldest ages. However, eliminating all individuals with fewer than fifteen features would effectively cut the sample size in half (see Figure 6.1) and remove a large number of individuals that appear to be aged effectively using the new procedure. For these reasons, both archaeological samples are

³⁸ A procedure known as Calibrated Expert Inference (CEI) that integrates experience-based estimates into a statically valid age-estimation procedure has been under development by other members of the research team for many years (Boldsen, 2009; Weise, Boldsen, Gampe, & Milner, 2012). This procedure, however, requires each user to produce a sample of experience-based age estimates from a known-age skeletal sample, which are then operationalized into a user-specific version of the method. Therefore, this technique is not practical for the majority of the field.

analyzed here using the full set of 40 possible features. Any individual with fewer than five traits present will be removed from the sample used to construct mortality distributions.

One difference in Figure 6.2 that should be addressed between the estimates of the St. Bride's and HOM 1272 samples is that a greater number of archaeological individuals (lower row) fall above the identify line. A linear regression fit through the origin for these data would have a steeper slope than for the historical sample (top row). Although individuals of the HOM 1272 sample could potentially be aging "faster" than those from St. Bride's (i.e, older traits appearing at younger ages), there are other potential explanations similar to those discussed in Chapter 5 regarding differences between the Athens and St. Bride's samples. The tendency for the uniform prior distribution to result in overestimates of age in the upper half of the lifespan, paired with fewer traits present for analysis, may be contributing to this pattern. Slight underestimations of age at the upper end of the age distribution in the experience-based estimates, which more often occur when few traits are available, could magnify this effect.

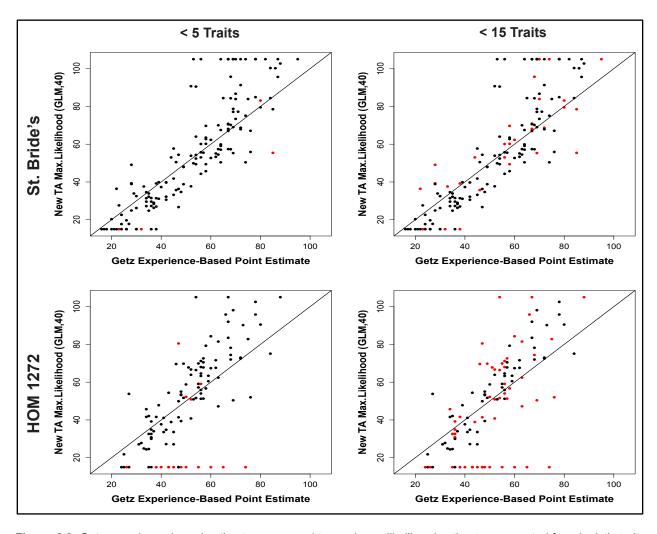


Figure 6.2. Getz experience-based estimates compared to maximum likelihood estimates generated from logistic trait probabilities for a maximum of 40 features for the St. Bride's sample (N=168) and Klosterkirke (HOM 1272, N=166). The same data are shown in the panels for each sample; estimates produced using fewer than 5 or 15 features are shown in red.

Preservation of commonly used adult age indicators

As with the traits used for the new TA method, the preservation of commonly used age indicators is a potential biasing factor in any skeletal analysis. Figure 6.3 shows the percentage of each archaeological sample that could be aged using each method. As was seen in the St. Bride's crypt sample, the pubic symphysis was less well preserved than the auricular surface in both sets of archaeological remains. Although the pubic symphysis is the most widely used adult age indicator, and considered by many to provide the most accurate age estimates, it is rarely preserved in archaeological samples (M. Cox, 2000; Falys & Lewis, 2011; Garvin & Passalacqua, 2012). This is not entirely unexpected because the pubic symphysis has relatively thin cortical bone, paired with an anterior position in the skeleton, which makes the area particularly prone to weathering and other postmortem damage (M. Cox, 2000).

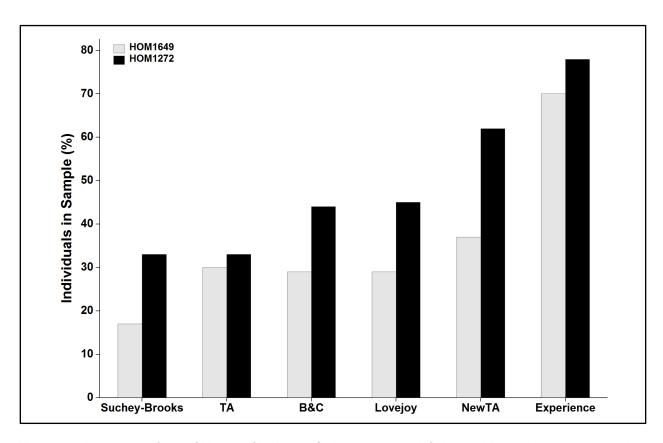


Figure 6.3. Percentage of the HOM 1649 (Ole Worms Gade, N=317) and HOM 1272 (Klosterkirke, N=166) samples that could be aged using methods for the pubic symphysis (Suchey-Brooks), auricular surface (Lovejoy and B&C), and those that use features from multiple skeletal elements (experience, existing TA, and new TA).

The performance of standard age indicators on past populations

It is commonly believed that rates of aging in skeletal indicators in the past may be different than they are in modern populations, although whether the change is faster, slower, or perhaps simply obscured by excessive "wear and tear" on the skeleton is debated (Chamberlain, 2006; Milner et al., 2008; Molleson, 1995). Therefore, before applying standard age-estimation methods to the archaeological samples, whether the relationship between traditional methods and experienced-based estimates is the same in both the historical and archaeological samples is explored. As previously noted, although experience-based estimates cannot provide quantitative measures of uncertainty, they have been shown to be unbiased and highly correlated with documented age in more than a half-dozen skeletal samples. Figure 6.4 shows the relationship between experience-based estimates of age and the ranges produced for two standard methods, existing TA, and the new TA procedure using 40 features. Although the archaeological sample (Figure 6.4, right column) has many fewer individuals who could be scored using standard techniques, the overall relationship to the experience-based age estimates, as well as the patterns inherent to each method are the same in both samples. The increase in variation—longer and more variable range lengths—seen in the existing and new TA methods for the archaeological sample is the result of poorer preservation compared to the individuals in St. Bride's crypt.

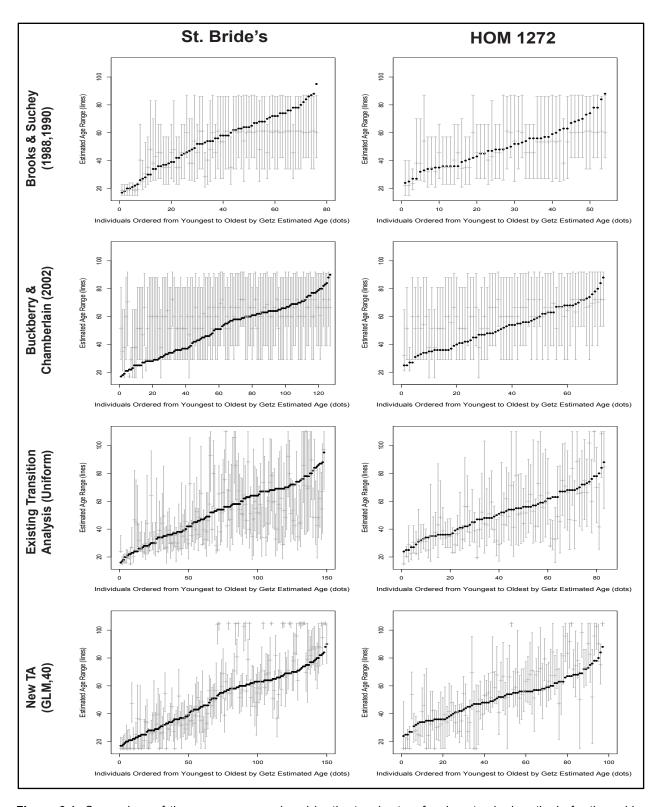


Figure 6.4. Comparison of the age ranges produced by the two best-performing standard methods for the pubic symphysis and auricular surface, existing TA, and the new TA procedure using 40 features for the documented St. Bride's sample and the Klosterkirke archaeological sample. Black dots in each graph show Getz experience-based estimates, while the grey lines indicate the range estimated for these individuals by each method.

Standard Methods & Implications for Paleodemographic Applications

As previously discussed, for paleodemographic analyses, whether the data adequately represent the sample is a fundamentally different problem than if the sample can be used to investigate the population. Although both questions are critical to the results and interpretations of archaeological investigations, if the input data are an inadequate representation of the sample, the application of any statistical analysis or modeling technique to generate information about the population will produce meaningless results. Therefore, the remainder of this dissertation focuses on the issue of whether data from any of the techniques investigated can produce results suitable for more sophisticated analyses.

Three standard age-estimation methods

This phase focuses on three of the most commonly used procedures applied to archaeological data—Suchey-Brooks for the pubic symphysis and Lovejoy et al. and Buckberry-Chamberlain for the auricular surface (Falys & Lewis, 2011; Garvin & Passalacqua, 2012). Although a variety of other techniques could be evaluated in a similar fashion, these were selected to be representative of the types of methods that are currently in use. Furthermore, their extensive use for bioarchaeological and paleodemographic investigations, particularly for the Suchey-Brooks and Lovejoy procedures, makes the critical evaluation of these techniques using modern, historical, and archaeological samples all the more meaningful.

The six-phase Suchey-Brooks method (see Figure 2.0) is based on a sample of 1,225 pubic symphyses collected from autopsies in Los Angeles (Brooks & Suchey, 1990). The authors state that this diverse sample should represent the variation found in prehistoric populations³⁹ and that "for prehistoric reconstruction there is no legal framework and age identification can focus on the mean values," (Brooks & Suchey, 1990, p. 237). Additionally, the method should be useful in archaeological contexts because the pubic symphysis is relatively often preserved and the mean values for each for the six stages are similar for males and females, which may simply the construction of mortality profiles (Brooks & Suchey, 1990).

The Lovejoy et al. (1985) eight-phase auricular surface technique is also based on a large sample of 500 known-age skeletons from the Todd Collection, several forensic cases, and 250 prehistoric individuals, which were presumably used to investigate the range of variation seen in the feature in archaeological skeletons. The authors state that the auricular surface produces age-at-death estimates that are similar in accuracy to those produced by the pubic symphysis, but that are significantly more useful because the features often better preserved (Lovejoy, Meindl, Pryzbeck, et al., 1985).

The Buckberry and Chamberlain (2002) method is based on the features of the Lovejoy et al. (1985) method, but is fundamentally different from that method in several respects. First, the features are scored as components, rather than in phases, which allows partial remains to be scored, as well as results in a greater number of possible age estimates. Secondly, the sample used to develop the method was a small group of individuals from the Spitalfields Collection (N=180)—a historical collection formed from individuals who lived contemporaneously in London with the individuals in St. Bride's crypt. Therefore, based on the idea that population-specific methods may produce better results, the performance of this method can be compared between the St. Bride's crypt sample and the other samples composed of individuals who lived substantially different lives than those from nineteenth century London.

_

³⁹ This view is in line with the conservative approach advocated by the research team.

These methods were applied according to the procedures described in the original publications. For consistency the left side was used whenever possible. If the left was unable to be scored, the right side was substituted. In cases where the pubic symphysis or auricular surface was only slightly damaged, multiple scores were assigned, the ranges were combined, and the points of central tendency were averaged. As mentioned in Chapter 5, the mean age estimates from the stages of existing age-estimation methods are the most frequently used data for generating population age-at-death distributions from archaeological data (Hoppa & Saunders, 1998; Jackes, 2000).

First, the mortality distributions produced by each method for the St. Bride's sample are compared to the known age-at-death distribution of the same individuals (Figure 6.5, top row). The estimated distribution for St. Bride's sample is then compared to the distribution estimated from the Klosterkirke (HOM 1272) archaeological sample (Figure 6.5, bottom row). To standardize this comparison across methods and samples, the individuals in each five-year age category are represented as a percent of the entire sample. The overall shape of the distributions is the most important feature in these comparisons. None of the age-at-death distributions generated using traditional methods successfully approximate the known age distribution of the St. Bride's sample (top row), nor do they resemble each other. Surprisingly, however, the distributions estimated for the St. Bride's and archaeological samples are almost identical for each method (Figure 6.5, bottom row). In other words, the mortality distributions produced for these samples appear to reflect the method that was used rather than the age distribution of the samples.

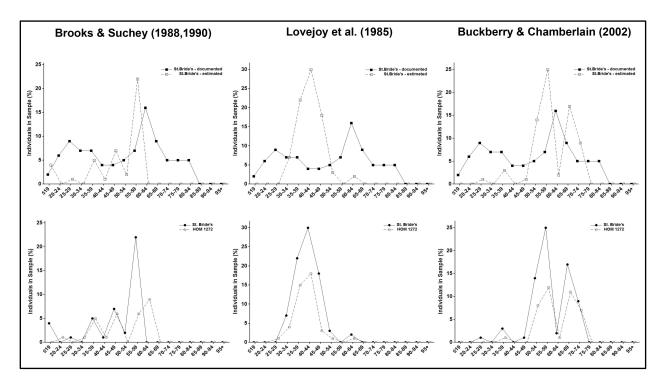


Figure 6.5. Top row: The documented (solid line) and estimated (dashed line) age-at-death distributions for the St. Bride's sample do not resemble one another. Bottom row: Estimated age distributions for the St. Bride's and HOM 1272 (Klosterkirke) samples show similar method-specific patterns.

Without knowing that the estimated distribution for the St. Bride's sample does not approximate reality, it would perhaps be tempting to make inferences about the similarities between the two groups that resulted in similar mortality profiles. This may be particularly true in this case because the two samples lived in similar environments, under similar social conditions, during the same time period. This scenario becomes even more likely because researchers often favor one method over another and rarely directly compare mortality distributions generated from different features independently, such as is done here.

The picture becomes even more troubling when the mortality distributions generated for the other archaeological sample (Ole Worms Gade, HOM 1649) and the modern individuals from the Athens sample to this comparison (Figure 6.6). Not only are the known differences between the mortality profiles of the Athens and St. Bride's samples (see Figure 5.0) not apparent in this comparison, but each method produces an extremely similar age distribution for each sample. This pattern is all the more remarkable because of the differences among the samples in terms of the diet, health status, and activity level of the individuals who formed them, combined with the differences in how the samples were formed and their level of preservation. In other words, the tendency of these commonly used procedures to produce method-specific patterns seems to be robust, and is not simply the by-product of analyzing four nearly identical samples.

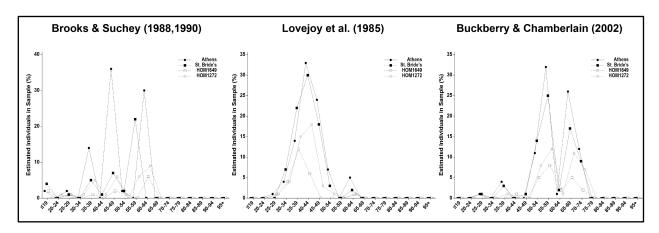


Figure 6.6. Comparison of the mortality distributions produced for the Phase 3 know-age samples (Athens and St. Bride's) and Phase 4 archaeological samples (HOM1649 and HOM 1672) based on age-estimates from three standard methods.

These results support the conclusion from Phase 3 that existing techniques do not, and fundamentally cannot, estimate age with sufficient accuracy and precision to produce reasonable age-at-death distributions. The clear and consistent method-specific patterns found in all four of the samples are consistent with what is reported in the literature. Samples assessed using the Suchey-Brooks and Lovejoy et al. techniques show few individuals surviving beyond age 60, while the Buckberry-Chamberlain method is said to perform better on samples with an older age distribution (Chamberlain, 2006; Falys, Schutkowski, & Weston, 2006; Hens, Rastelli, & Belcastro, 2008; Nagaoka & Hirata, 2008; Pfeiffer, 1985; San Millán, Rissech, & Turbón, 2013). The data in Figure 6.6 also provide strong support for the long-discussed idea that the oddities seen in many archaeological age-at-death distributions are primarily the result of peculiarities in the methods used to evaluate the samples (Bocquet-Appel & Masset, 1982; Howell, 1982; Konigsberg & Frankenberg, 1992).

Existing TA, New TA, and experienced-based estimates

The Phase 3 validation study strongly indicates that the new TA method is capable of producing more realistic mortality distributions than traditional techniques. The mortality distributions produced for the St. Bride's, Athens, and two archaeological samples using existing TA, observer experience and new TA are presented in Figure 6.7. Although the patterns in this figure are less clear than those of standard methods, it is precisely this complexity that provides potentially valuable information.

The overall pattern produced by existing TA is similar to what was seen in the Phase 3 validation study with age being slightly underestimated before the age of 50 (i.e., a younger peak in the age distribution). The peak of adult mortality is slightly older in the experience-based estimates, which are more similar to those produced by the new TA method than other techniques. The differences between experience-based estimates and new TA in this comparison are exaggerated because of the previously noted issue of heaping at the ends of the age distribution. If these individuals are removed, the similarities between the two distributions become more apparent. Most importantly, all three of these methods show some indication of differences in the number of individuals who lived beyond 50 years of age in each of the samples. These differences were not, and cannot be, detected by standard methods.

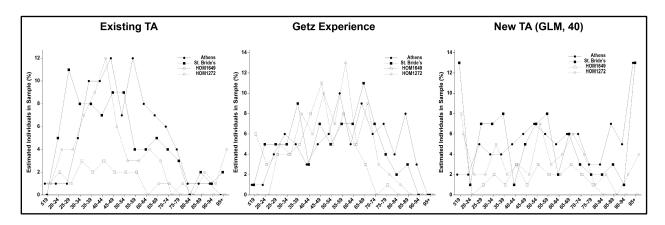


Figure 6.7. Comparison of the mortality distributions produced for the two Phase 3 known-age collections (Athens and St. Bride's) and the two Phase 4 archaeological samples (HOM1649 and HOM 1672) using existing TA, Getz experience-based estimates, and new TA.

Assessing Mortality in the Danish City of Horsens

Based on the analyses presented here and in Phase 3, it appears that, for the first time, there is now a method capable of producing statistically valid age estimates for the entire adult lifespan using even fragmented remains. Therefore, the final step in this application is to compare the mortality profiles produced by the various procedures to assess what they may indicate about changes in mortality in the late Middle Ages and Early Modern period. Figure 6.8 compares the age-at-death distributions produced for the two Danish archaeological samples (HOM 1649 and HOM 1272) using the three standard methods discussed above (left column), existing TA, experience-based estimates, and the new TA method (right column).

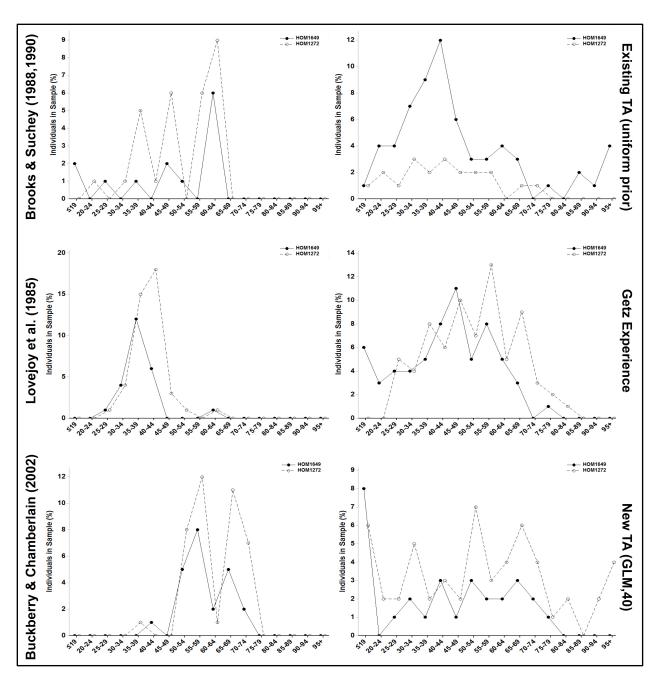


Figure 6.8. Age-at-death distributions for the individuals of the Ole Worms Gade (HOM 1649, solid line) and Klosterkirke (HOM 1272, dashed line) samples based on estimates from commonly used techniques (left column), existing TA, observer experience, and new TA (right column).

As previously discussed, the standard methods produce essentially the same age distribution for both samples (Figure 6.8, left column). The larger peaks in the age distributions in the Klosterkirke sample (HOM 1272, dashed lines) for most methods is simply the result of more individuals having the appropriate elements available for analysis; it is the overall shape of the distribution that is key to interpreting differences between the samples. The only significant deviation between the two samples among the three traditional techniques is that, for Ole Worms Gade (HOM 1649, solid line), the Suchey-Brooks technique was able to detect a

greater number of individuals under the age of 20. Based on experience-based age assessments (Figure 6.8, right middle), this elevation in mortality in young adulthood for the earlier of the two sites is a real feature of the sample and not merely an error of the method. Any excitement should be tempered, however, by the fact that for individuals in their teens and early twenties there are a number of other indicators that provide more precise information than the pubic symphysis. So although the method may have detected an important demographic difference between the samples, the use of other age indicators in conjunction with these data would provide a more precise picture.

Existing TA shows a different pattern for both of samples compared to the traditional methods (Figure 6.8, upper right). There is a large spike in mortality in the early forties for the Ole Worms Gade sample (solid line), while the pattern for Klosterkirke (dashed line) is similar to that of the St. Bride's sample. A naive interpretation of this difference might be that the Klosterkirke sample is more similar to the St. Bride's sample because the individuals lived during the same time period and shared similar cultural and environmental conditions. Upon closer inspection, however, an alternative emerges. While slightly more than 40% of the St. Bride's sample had relatively intact pubic symphyses, less than 15% of the individuals from Ole Worms Gade had this feature. This places a heavy emphasis in the existing TA estimates on the more robust features of the auricular surface, which were originally developed based on concepts from the Lovejoy et al. (1985) descriptions. Based on this, it is not perhaps not surprising to see a large peak in mortality in the same age range as the Lovejoy et al. method.

By far, the most interesting distributions are those produced using experienced-based estimates and the new TA procedure. Although somewhat muddled by the errors at the extremes of the distribution, which were discussed in previous sections, the estimates produced by new TA show several of the same features as the distribution produced from experience-based estimates. Most notably, both distributions show similar peaks in middle-age between 35 and 75 years of age and, contrary to what is found using standard procedures, many individuals living beyond the age of 50. In Figure 6.8, both experience-based estimates (right middle) and new TA (bottom right) also show a slightly older age distribution in the Klosterkirke sample (dashed lines).

Producing similar sex-separated age distributions for these archaeological samples results in more complex patterns because of the extremely small sample size in each five-year age category. Comparisons between the distributions are further hindered by differences in the number of males and females in each sample, differences in preservation between the sexes and samples, and the heaping of individuals that currently occurs in the youngest and oldest age categories. More sophisticated analytical and statistical procedures are necessary to deal with these issues to confidently evaluate sex-specific differences in mortality.

Conclusions

At a glance, Figure 6.8 reveals significant variations in the data produced by each of the six methods evaluated in this application. Based on the vast differences in the mortality distributions, there is little doubt that summarizing these data into tables or analyzing them with more sophisticated techniques would also produce different results. In this dissertation, the application of these methods by a single observer to both modern and historical skeletons prior to applying them to archaeological samples provides critical context for the interpretation of the age-at-death distributions. Independent of the test data from Phase 3, it would be impossible to know which of the distributions was likely to best represent the true age structure of the mortality sample.

This first application of the new TA procedure to archaeological samples provides an exciting glimpse of possible differences in mortality in the city of Horsens between the late Middle Ages and Early Modern period that could not previously be detected. In the distributions based on Getz experience and new TA, there was young adult mortality in the Ole Worms Gade sample from the late Middle Ages and an increased number of individuals dying at older ages in the Klosterkirke sample from the Early Modern period. Although this pattern is precisely what we would expect to see based on historical documentation, these results offer only preliminary indications of potential differences between these samples. In addition to the error introduced into the new TA estimates by using a uniform prior distribution and not including young-age features of the pelvic joints, there are indications that the samples from the two sites may not be representative of their respective populations. Recall that the Ole Worms Gade sample potentially represents less than 10% of the entire cemetery and that fewer than half of these individuals could be evaluated using any of the methods, including new TA.

When the additional reference data collected as part of the NIJ project are available, new mortality profiles will be produced for both of the archaeological samples using a larger suite of traits. This should increase the sample size avilable for analysis, as well as provide more accurate estimates for the individuals already assessed in this application. The new features added as part of the NIJ project, combined with additional statistical improvements to the preliminary TA procedure, may also help to resolve the age-estimation issues at the extreme ends of the age distribution. It is hoped that with larger sample sizes, differences in sex-specific mortality can be better addressed. Additionally, a more in-depth assessment of the context of the samples will be used to help bridge the gap between the analysis of archaeological samples to making interpretations about the populations that they represent.

Although the new TA method is still in a preliminary form, it is clear that pairing the statistical ideas of the existing TA procedure with a large number of age-informative features has been a success. More realistic approximations of age can be produced than ever before, with the prospect of even better estimates in the not so distant future.

CHAPTER 7: DISCUSSION, CONCLUSIONS, & FUTURE DIRECTIONS

This dissertation, in conjunction with the work of the research team, is the first large-scale attempt to produce accurate, precise, and unbiased age estimates for the entire adult human lifespan using features throughout the skeleton. Figure 7.0 provides a timeline of the research conducted as part of this dissertation and the collaborative work performed as part of the NIJ project. Table 7.0 summarizes the samples used in each phase of this work.

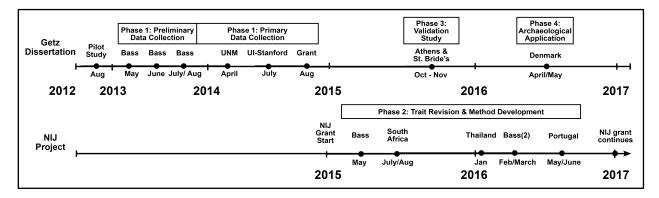


Figure 7.0: Timeline for this dissertation and the collaborative work conducted as part of the NIJ project.

Building on over two decades of preliminary work by the research team, in Phase 1 more than 200 trait variations were investigated to identify and refine a set of age-related features in five preliminary rounds of data collection. Fifty-three traits were then evaluated in a sample of 1,010 individuals from four known-age North American skeletal collections to form the primary reference for later analyses. Based on this combined sample, logistic and generalized additive models were used to identify forty-five features with potentially useful age-related variation.

These features were included in the more extensive list investigated by the NIJ-funded research team using skeletons of differing ancestry from four continents. The NIJ project and its relationship to this dissertation were described in Chapter 4 (Phase 2). In particular, two members of the research team (Milner and Boldsen), made substantive contributions to this dissertation in the areas of trait identification and refinement, and the statistical methods used to analyze this information, respectively. Discussion with all members of the research team related to the traits and the nature of aging are reflected in the excerpts of the draft NIJ trait manual in Appendix D, and throughout this dissertation.

In Phase 3, a simplified version of the existing TA procedure was put into practice using the reference data collected in Phase 1 and two independent, known-age European samples. The Athens Collection, a sample of modern well-preserved, individuals who lived the majority of their lives after WWII, was used to evaluate the method's performance under essentially ideal conditions. Over 70 variations of the new TA procedure were tested using combinations of different features, statistical models, reference samples, and prediction intervals to identify the effects of each choice on the accuracy and precision of the estimates produced. The possible effects of each factor were considered and the best-performing combinations were selected for additional testing on historical and archaeological skeletons.

Table 7.0. Samples collected in each research phase

Research Phase	Collection	Location	N	M	F
Phase 1a: Preliminary Trait Identification & Refinement	WM Bass	Knoxville, TN	124	66	58
	WM Bass	Knoxville, TN	122	62	60
	WM Bass	Knoxville, TN	101	55	57
	WM Bass	Knoxville, TN	196	91	105
Phase 1b: Primary Trait Selection	WM Bass	Knoxville, TN	500	272	228
	Maxwell Museum	Albuquerque, NM	170	101	69
	UI-Stanford	Iowa City, IA	149	129	20
	JCB Grant	Toronto, ON	191	175	116
Phase 2: NIJ Trait Refinement & Method Development	WM Bass	Knoxville, TN	423	221	202
	Pretoria	Pretoria, ZA	424	269	155
	Chiang Mai	Chiang Mai, TH	440	271	169
	WM Bass (2)	Knoxville, TN	14	12	2
	Lisbon	Lisbon, Portugal	390	190	200
Phase 3: Method Validation	St. Bride's Crypt	London, UK	168	92	78
	Athens Donated	Athens, GR	201	111	90
Phase 4: Archaeological Application	Ole Worms Gade	Odense, DK	317*	158	129
	Klosterkirke	Odense, DK	166*	85	78

^{*} Total number of individuals in the Phase 4 samples is not equal to the combined number of males and females because sex could not be estimated for every individual.

The simplest three statistical model tested—the logistic curves—based on a combined, non-sex-specific reference sample produced the best compromise between accuracy and precision. Age estimates produced using 30 features were slightly more accurate than those produced using all 40 possible traits, possibly because of the inappropriate narrowing of age intervals caused by highly correlated features. Estimating age with smaller feature sets—15 features based on expert judgment and 21 iteratively selected features—resulted in slightly wider age intervals with correspondingly higher accuracy; however, using fewer possible features at the outset greatly reduces the utility of the new TA method for partial remains. For the Athens sample, estimating age using trait probabilities from the modern individuals in the post-1981 death cohort did not improve method performance. However, age estimates produced using the pre-1953 death cohort for the same individuals resulting in the highly atypical combination of wider age ranges with lower accuracy.

The new TA method was then applied to a less well-preserved, known-age sample formed from a population a step further back in time—a historical sample from St. Bride's Crypt in London, England—from the same time period as one of the Phase 4 archaeological samples. The overall accuracy, precision, or bias of the age estimates produced for the Athens and St. Bride's samples were essentially the same, despite differences in the diet and health status of the individuals in each collection. As with the Athens sample, using a time period-specific reference sample did not improve the collective accuracy or precision of the age estimates. Surprising

however, unlike the Athens sample, estimating age based on probabilities from modern individuals (post-1981 death cohort) also had little effect on the performance of the method.

Finally, in Phase 4, the new TA procedure was applied to two archaeological samples selected to represent the late Middle Ages and Early Modern periods in Denmark—Ole Worms Gade and Klosterkirke—where historical records indicate that changes in mortality patterns are likely to have occurred. Although standard methods of age estimation were hindered by the loss or damage of key portions of the pelvis, the new TA procedure was able to produce age estimates for nearly twice as many individuals in both archaeological samples. The three standard methods for the pubic symphysis and auricular surface were shown to produce essentially the same method-specific mortality distributions for the Athens, St. Bride's and Danish samples despite known differences among these collections. Because none of the estimated distributions produced using standard techniques approximate the true age distribution of the known-age samples, the representations of past mortality are also highly suspect. Until additional testing is conducted, the age-at-death distributions produced for all past populations using these techniques should be considered, at best, questionable, and, at worst, fiction based on inherently flawed methods. To emphasize this point, if the patterns of age-at-death produced by each of the standard methods are truly universal, a researcher would only need to select a method to predict the mortality profile for a population—no evaluation of skeletons would be required.

Discussion

Trait selection and data collection procedures

The traits used in this dissertation were refined by Milner and Getz in Phase 1 with some input from other members of the NIJ research team, but the data used in Phases 1, 3 and 4 were collected by a single individual. Because the vast majority of research shows that intra-observer error is typically lower than inter-observer error, it is possible that the new TA method may perform slightly better here than when applied by others; however, this possibility is mitigated by two factors. First, the reference data and those for the validation study and archaeological applications were collected over a period of several years. Data collection trips for this dissertation were interspersed with intensive trait refinement and work with diverse skeletal collections in collaboration with the NIJ research team. These factors likely introduced subtle shifts in trait definitions over time, particularly between Phases 1 and 4. These potential differences between the reference data and test samples introduce an unknown level of error into the age estimates that were produced.

In this dissertation, no features from the two most widely used skeletal elements for adult age estimation—the pubic symphysis and auricular surface—were included in the new TA procedure. Although the inclusion of some pelvic joint features would undoubtedly improve age estimates, particularly in young adulthood, excluding these features allowed the value of previously unknown or under-utilized skeletal features to be evaluated. Importantly, this work demonstrates the collective strength of low information traits throughout the skeleton evaluated in a TA framework because it is possible to produce accurate and precise estimates throughout adulthood without the inclusion of traditional features. That said, modifications to some of the existing pubic symphysis and auricular surface features of existing TA have been made by the NIJ research team and will be included in the final version of the new TA procedure released to the osteological community. Additionally, because initial work on the new TA procedure focused on identifying features that provided age-related information in middle and old age, very few features with transitions in young adulthood were included in the tests presented here. Because

of this, individuals under 30 years of age showed higher error than would be expected in this portion of the lifespan and had a tendency to be placed into the youngest age category in the mortality distributions. This issue will be corrected as part of the NIJ project by adding additional features, such as fusion of vertebral epiphyses and iliac crests, which provide critical information in this age range. These additional data were collected for individuals in the Athens and St. Bride's samples and can be included in future analyses to evaluate their effect on method performance.

Correlated traits

In the variations of the new TA method tested here, only features from the left side of the skeleton were used. Although there are often bilateral differences in trait expression, relatively high correlations between bilateral features likely exist. Using data from only one side of the body reduces the potential error that could be introduced by including additional correlated features and eliminates the variability in age estimates that could be introduced by right-left asymmetries in trait age distributions. Differences between trait development in the right and left sides of the skeleton may result from important activity-related differences, but they may also be heavily influenced by the elements present for evaluation in the reference samples. The decision to use data from only one side of the skeleton does, however, have practical consequences for both the number of individuals that can be evaluated and, potentially, the quality of the age estimates produced. Because the right side was not used if the left was missing or damaged, fewer skeletons could be evaluated, particularly from the archaeological samples. Additionally, as shown in Chapter 5, the accuracy and precision of age estimates tends to increase, to a point, when a greater number of features are present. Therefore, age estimates may improve for some individuals in the St. Bride's and archaeological samples, when traits from the right side are included, particularly when the left side is unavailable.

An evaluation of the Phase 1 reference data set presented in Chapter 3 and Appendix E revealed a number of highly correlated trait pairs; however, no statistical correction has been included in the tests of the new TA procedure presented here. Using a reduced set of features selected by choosing the most age-informative feature from each highly correlated pair is a practical method that showed potential for reducing the effects of correlated traits. Unfortunately, the accuracy of the ages produced using this method variation still fell short of what would be expected using a 95% prediction interval. Using fewer traits also appears to increase variability in the point estimates produced and results in a potentially significant loss of useful data, particularly when only partial remains are available for analysis.

Existing TA currently includes a statistical correction to reduce the influence of correlated traits on the estimated age ranges. However, tests of the method have shown that overall accuracy in several samples is still lower than would be expected using 95% prediction intervals (e.g., Figure 5.1), indicating that the effect of correlated features has not been entirely eliminated. Although it is still a work in progress by the NIJ research team, the final version of the new TA program released for general use will likely include a combination of both practical and statistical solutions. Where neither approach has succeeded independently, simultaneously selecting the most age-informative set of features with the lowest correlations and including a statistical correction may result in a better compromise between accuracy and precision than can be currently obtained.

Sex- and population-specific reference standards

Although conventional wisdom suggests that models generated using samples of individuals who are more similar to those being evaluated would generate more accurate and precise results, this was not seen in the modern or historical collections evaluated in Phase 3. While the use of appropriate time period-specific probabilities does not result in a great improvement, using an inappropriate reference sample appears to have a significant adverse effect, at least for modern individuals. This is potentially the result of the large differences in sample size between the death cohorts. The lack of females in the earlier death cohort also prevented sexspecific models for the historical St. Bride's sample to be evaluated. Significantly larger samples of known-age individuals born in the nineteenth century would be needed to effectively evaluate the effects of using time period-specific reference samples and to confidently say something about secular changes in aging during the past several centuries.

Prior distributions

The versions of the new TA method tested in this dissertation assume a uniform prior distribution of age-at-death where all individuals between 15 and 105 years of age are equally likely to end up in a mortality sample. The uniform prior distribution can be considered a conservative approach because it does not impose a specific age-at-death structure onto the data. However, it does impose unrealistic assumptions that have been shown to contribute to overestimates of age in older adults using the existing TA procedure. Fortunately, tests of existing TA with the Bass Collection (Milner & Boldsen, 2012c) and the Athens sample (Figure 4.16) have shown that the choice of prior distribution makes very little difference until after around age 60 and no substantial difference until after age 80. In the archaeological applications, the new TA procedure was compared to the results generated by existing TA also using a uniform prior. So, although the existing TA method includes a correction for correlated features, the differences between existing TA and new TA are essentially only the result of the features used to generate the age estimates.

Conclusions

New statistical procedures are an important component in the process of moving the field towards more quantitatively justified age estimates. However, their application to poor skeletal indicators cannot possibly resolve many of the fundamental aspects of the adult age-estimation problem. Although differences in the aging process between the sexes and among populations have important implications for the study of the human species, existing methods of age estimation are too blunt a tool to contribute to these investigations. In this dissertation, several of the most commonly used methods in bioarchaeological and paleodemographic studies where applied to modern, historical, and archaeological samples. For the St. Bride's sample, the mortality distributions produced using three traditional methods neither resembled the true age distribution of the sample, nor each other. It was then shown that almost identical methodspecific patterns were produced in the mortality distributions for the known-age Athens sample and two Phase 4 archaeological samples. This direct comparison of the same methods, applied by the same observer to four different samples, revealed that each of the traditional methods tested produced a characteristic pattern of adult mortality that was essentially unaffected by differences in the age distribution of the sample evaluated. This clearly demonstrates that the need for population- and sex-specific refinements of existing methods pales in comparison to the importance of finding new age-informative traits that provide information throughout adulthood and creating new procedures for using this information effectively

This dissertation is a large-scale proof-of-concept that accurate and reasonably precise age estimates without significant bias can be produced for all of adulthood without using either of the most two most commonly used age indicators—the pubic symphysis and auricular surface. The new TA method produces a similar level of accuracy to procedures widely used by archaeologists and forensic anthropologists—Suchey-Brooks (1988,1990) and Buckberry-Chamberlain (2002)—but with point estimates that strongly correlate with documented age and intervals that are, on average, half as wide. These results were obtained despite using a relatively simple logistic model to generate non-sex-specific trait probabilities from a diverse reference sample, and combining this information without correcting for correlated traits or using an informative prior distribution.

Using the new TA method, there is reason to be optimistic about the future of paleodemography. Based on the results of Phases 3 and 4, we can be confident that the ability to produce accurate mortality profiles for past populations is closer than ever before. Although refinements are needed to the existing procedure to correct several issues and make it widely applicable to other populations, we have, for the first time, a method capable of producing statistically valid age estimates for the entire adult lifespan using even fragmented remains.

Ongoing Work

That NIJ-funded research team project is scheduled for completion in late 2017, with the release of a computer program and full scoring manual in early 2018. This project addresses a number of issues that could not be fully explored as part of this dissertation, including: 1) the selection of appropriate statistical models (GLM and GAM variations), 2) additional methods for identifying correlated traits and mitigating their effects, 3) evaluating inter- and intra-observer error, 4) assessing population variation related to ancestry, 5) the efficacy of different informative prior distributions, and 6) the development of computer software and materials for dissemination of the new method. At the conclusion of the NIJ project, a new version of the TA program—an elborated version of the technique used in Phases 3 and 4—will be disseminated to the entire osteological community as both a free, downloadable, stand-alone program and as part of an already widely used software package—Fordisc (Ousley and Jantz 2005). Trait definitions, diagrams, and photographs created in Phases 1, 2, and 3 will be included as part of both the scoring manual and computer software.

Data for the new features identified and refined by the NIJ team were collected for the Athens, St. Bride's, and Danish archaeological samples. The larger NIJ data set will be used to reevaluate the performance of the new TA method for these samples to assess the effect of adding additional features. The effects of including more features that provide additional information, particularly in very young and old age, as well as using a broader, more diverse reference sample will be assessed. Correlations between and among features will continue to be assessed, including those between the right and left sides of the body and those in areas of the skeleton that may be biologically or functionally related. New mortality profiles will also be produced for the two Danish archaeological samples using additional data from individuals of European ancestry collected as part of the NIJ project. These new age-at-death distributions will be evaluated in light of additional contextural information about the samples and demographic changes known to have occurred in the late Middle Ages and Early Modern period as a result of dietary changes, increased mobility, and improved sanitary conditions.

Future Directions

Research related to the osteological aspects of this project will continue on multiple inter-related fronts. These include the identification of new traits and expansion of the existing reference data set using additional known-age skeletal collections. Variations of the new procedure, including statistical and methodological modifications incorporated by the research team, will be tested on combinations of the samples from the NIJ project and this dissertation, as well as on new samples as they become available. Future work will also investigate how forms of two- and three-dimensional data, including photographs, videos, laser scans, and computed tomography (CT), can be used in the documentation and analysis of skeletal features. Documenting features using these technologies could allow new traits to be assessed without the need to return to collections, which is not only costly and often impractical, but potentially damaging to fragile skeletal remains. These data, archieved in publically available databases, would also represent an invaluable resource for researchers and the teaching of human variation. Future work will assess the feasibility of investigating patterns in age-related features using CT data from living and recently deceased individuals. If successful, this line of investigation will allow age-related variation to be assessed on a scale that is not possible with skeletal samples.

This dissertation, viewed in conjunction with the work being done by NIJ research team, provides every indication that a version of the new TA method may become the universal gold standard for adult age estimation.

APPENDIX A: PUBLICATIONS RELATED TO ADULT SKELETAL AGE ESTIMATION USING MACROSCOPIC FEATURES OF THE CRANIAL SUTURES, RIBS, PUBIC SYMPHYSIS, & SACRO-ILIAC JOINT

This appendix is an extensive, but not exhaustive, list of publications related to age-at-death estimation for adult skeletons using the cranial sutures, ribs, pubic symphyses, sacro-iliac joint, and combinations of these features. Age-estimation procedures that have been tested, applied, or revised are listed individually with their related publications by category (e.g., tests/applications or revisions/modifications). Publications that both test an existing procedure and present revisions or test multiple procedures are listed multiple times as appropriate. Publications presenting general age-related observations on a particular part of the skeleton and methods that have not yet been evaluated by other researchers are grouped together in a single list (observations/other) at the end of each section.

Research dealing with histological features of bones or teeth, or traits that can only be evaluated from MRI or CT data are included only if they also test or apply macroscopic methods. The multitude of conference presentations, Master's theses, and PhD dissertations related adult skeletal age estimation are excluded from this list. An exception was made, however, for two theses, (Baker, 1984; Masset, 1982) dealing with the cranial sutures that are commonly cited in relevant publications and reference manuals.

Cranial Sutures

(Todd & Lyon, 1924, 1925a, 1925b)

Tests/applications: (Brooks, 1955; Dokladal, 1975; Meindl & Lovejoy, 1989; Singer, 1953)

(Acsadi & Nemeskeri, 1970)

Tests/applications: (Galera, Ubelaker, & Hayek, 1995; Khandare, Bhise, & Shinde, 2015; Wolff, Vas, Sótonyi, & Magyar, 2012)

(Masset, 1982)

Tests/applications: (Galera et al., 1995; Galera, Ubelaker, & Hayek, 1998)

(Baker, 1984)

Tests/applications: (Galera et al., 1995, 1998)

(Perizonius, 1984)

Tests/applications:(Harth et al., 2009, 2010; Key, Aiello, & Molleson, 1994; Lynnerup & Jacobsen, 2003; Sahni & Jit, 2005)

(Meindl & Lovejoy, 1985)

Tests/applications: (M. F. Anderson et al., 2010; Galera et al., 1995, 1998; Gocha et al., 2015; Hershkovitz et al., 1997; Key et al., 1994; Nagar & Hershkovitz, 2004; Saunders et al., 1992; Wolff et al., 2012)

Revisions/modifications: (Key et al., 1994; Nawrocki, 1998)

(Mann, Jantz, Bass, & Willey, 1991; Mann, Symes, & Bass, 1987)

Tests/applications: (Apostolidou et al., 2011; Ginter, 2005; Gruspier & Mullen, 1991)

Revisions/modifications: (Nawrocki, 1998)

Observations/other: (Abbie, 1950; Alesbury, Ubelaker, & Bernstein, 2013; Beauthier et al., 2010; Boyd, Villa, & Lynnerup, 2015; Broca, 1875; Chiba et al., 2013; Dorandeu et al., 2008; Dorandeu et al., 2009; Dwight, 1890a; Johnson, 1976; Kokich, 1976; Lynnerup & Jacobsen, 2003; Masset, 1971, 1989; N'Guyen, Gorse, & Vacher, 2007; Nagar & Hershkovitz, 2004; Obert et al., 2010; Parsons & Box, 1905; Powers, 1962; Saito, Shimizu, & Ooya, 2002; Sauvage, 1870; H. Schmitt & Tamáska, 1970; Singh, Oberoi, Gorea, & Kapila, 2004; Wenguang & Ke, 1989)

Ribs

(Michelson, 1934)

Tests/applications: (Garamendi, Landa, Botella, & Alemán, 2011)

(Işcan, 1991; İşcan & Loth, 1986a, 1986b; İşcan, Loth, & Wright, 1984a, 1984b, 1985; İşcan, Wright, & Loth, 1987; Loth & İşcan, 1989)

Tests and Applications: (Aktas, Koçak, Aktas, & Yemisçigil, 2004; Baccino et al., 1999; Cerezo-Román & Espinoza, 2014; Dedouit et al., 2008; Dudar, 1993; Dudar, Pfeiffer, & Saunders, 1993; Fanton, Gustin, Paultre, Schrag, & Malicier, 2010; Galera et al., 1995; Gocha et al., 2015; Gupta, Rai, Kalsey, & Gargi, 2007; Kimmerle, Prince, et al., 2008; Loth, 1995; Loth, İşcan, & Scheuerman, 1994; Martrille et al., 2007; Merritt, 2015; Nagar & Hershkovitz, 2004; Nikita, 2013; Oettlé & Steyn, 2000; Rejtarová, Hejna, Soukup, & Kuchař, 2009; Russell et al., 1993; Salem et al., 2014; Saunders et al., 1992; Telmon et al., 1996; Verzeletti, Cassina, Micheli, Conti, & De Ferrari, 2010; Verzeletti, Terlisio, & De Ferrari, 2013; Wolff et al., 2012; Yavuz, İşcan, & Çöloğlu, 1998; Yoder, Ubelaker, & Powell, 2001)

Revisions/modifications: (Hartnett, 2010b; Oettlé & Steyn, 2000; Salem et al., 2014; Verzeletti et al., 2010; Verzeletti et al., 2013)

(Kunos, Simpson, Russell, & Hershkovitz, 1999)

Tests/applications: (Kurki, 2005; Moskovitch et al., 2010; A. Schmitt & Murail, 2004)

Revisions/modifications: (DiGangi et al., 2009)

(Hartnett, 2010b)

Tests/applications: (Merritt, 2014)

Observations/other: (Barchilon et al., 1996; Davies, 1913; Elkeles, 1966; King, 1939; McCormick, 1980; McCormick & Stewart, 1988; Riebel, 1929; Semine & Damon, 1975; J. H. Stewart & McCormick, 1984)

Pubic Symphysis

(Todd, 1920, 1921, 1923)

Tests/applications: (Aiello & Molleson, 1993; M. F. Anderson et al., 2010; Bongiovanni, 2016; Brooks, 1955; Campanacho, Santos, & Cardoso, 2012; Galera et al., 1995; Jackes, 1985; Kimmerle, Konigsberg, Jantz, & Baraybar, 2008; Kimmerle, Prince, et al., 2008; Meindl & Lovejoy, 1989; Meindl, Lovejoy, Mensforth, & Walker, 1985; Molleson, 1995; Pal & Tamankar, 1983; Pfeiffer, 1980; Sinha & Gupta, 1995; Suchey, Wiseley, & Katz, 1986)

Revisions: (Brooks, 1955; Katz & Suchey, 1989; Meindl et al., 1985; Suchey et al., 1986)

(Gilbert & McKern, 1973; McKern & Stewart, 1957; T. D. Stewart, 1957)

Tests/applications: (Meindl et al., 1985; Pal & Tamankar, 1983; Suchey et al., 1986) (Aiello & Molleson, 1993; Angel, 1971; Galera et al., 1995; Gilbert, 1973; Houck et al., 1996; Johnston & Snow, 1961; Klepinger, Katz, Micozzi, & Carroll, 1992; Meindl & Lovejoy, 1989; Molleson, 1995; Pfeiffer, 1980, 1985; Sharma et al., 2008; Sinha & Gupta, 1995; Suchey, 1979)

Revisions/modifications: (Snow, 1983; Suchey et al., 1986)

(Acsadi & Nemeskeri, 1970)

Tests/applications: (Galera et al., 1995)

(Hanihara & Suzuki, 1978)

Tests/applications: (Chen, Zhang, & Tao, 2008; Meindl & Lovejoy, 1989; Sinha & Gupta, 1995; Zhaojin, 1988)

Revisions/modifications: (Zhaojin, 1988)

(Brooks & Suchey, 1990; Katz & Suchey, 1986, 1989; Suchey et al., 1988)

Tests/applications: (Baccino et al., 1999; Bednarek, Bloch-Bogusławska, & Sliwka, 2001; Berg, 2008; Blanc et al., 2005; Bongiovanni, 2016; Campanacho et al., 2012; Chen et al., 2008; Djurić, Djonić, Nikolić, Popović, & Marinković, 2007; Fleischman, 2013; Galera et al., 1995; Gauthier & Schutkowski, 2013; Gocha et al., 2015; Godde & Hens, 2012, 2015; Hens et al., 2008; Hoppa, 2000; Houck et al., 1996; Kimmerle, Konigsberg, et al., 2008; Kimmerle, Prince, et al., 2008; Klepinger et al., 1992; Lottering, MacGregor, Meredith, Alston, & Gregory, 2013; Martrille et al., 2007; Merritt, 2015; Nagar & Hershkovitz, 2004; Overbury, Cabo, Dirkmaat, & Symes, 2009; Pasquier et al., 1999; Rissech et al., 2012; Sakaue, 2006; San Millán et al., 2013; Saunders et al., 1992; A. Schmitt, 2004, 2008; Shirley & Ramirez Montes, 2015; Sitchon & Hoppa, 2005; Slice & Algee-Hewitt, 2015; Stoyanova, Algee-Hewitt, & Slice, 2015; Telmon et al., 2005; Telmon et al., 1996; Villa, Buckberry, Cattaneo, & Lynnerup, 2013; Wärmländer & Sholts, 2011; Wescott & Drew, 2015; Wink, 2014)

Revisions/modifications: (Berg, 2008; Corsini, Schmitt, & Bruzek, 2005; Godde & Hens, 2015; Hartnett, 2010a; Kimmerle, Konigsberg, et al., 2008; Konigsberg et al., 2008; Samworth & Gowland, 2007)

(Chen et al., 2008; Chen, Zhang, Zhu, & Tao, 2011) Tests/applications: (Fleischman, 2013)

(Hartnett, 2010a)

Tests/applications: (Merritt, 2014)

(Dudzik & Langley, 2015; Ferrant et al., 2009; A. Schmitt, 2008; Wang, Wang, & Tian, 2012; Zhongyao, 1986, 1995; Zhongyao & Yang, 1999; Zongyao, 1982)

Sacro-iliac Joint & Auricular Surface

Observations: (Resnick, Niwayama, & Goergen, 1975; Sashin, 1930)

(Bedford, Russell, & Lovejoy, 1989; Lovejoy, Meindl, Pryzbeck, et al., 1985)

Tests/applications: (M. F. Anderson et al., 2010; Barrier et al., 2009; Bongiovanni, 2016; Buckberry & Chamberlain, 2002; Galera et al., 1995; Gocha et al., 2015; Hens & Godde, 2016; Hens et al., 2008; Lovejoy, Meindl, Mensforth, et al., 1985; Maat, 1987; Martrille et al., 2007; Meindl & Lovejoy, 1989; Merritt, 2015; Mulhern & Jones, 2005; Murray & Murray, 1991; Nagaoka & Hirata, 2008; Rissech et al., 2012; Rougé-Maillart, Telmon, Rissech, Malgosa, & Rougé, 2004; Saunders et al., 1992; A. Schmitt, 2004, 2005; Storey, 2007; Telmon et al., 1996)

Revisions/modifications: (Buckberry & Chamberlain, 2002; Corsini et al., 2005; Osborne et al., 2004; Samworth & Gowland, 2007)

(A. Schmitt & Broqua, 2000)

Tests/applications: (Buk et al., 2012; Debono, Mafart, Guipert, & Jeusel, 2004)

(Buckberry & Chamberlain, 2002)

Tests/applications: (Bongiovanni, 2016; Falys et al., 2006; Gauthier & Schutkowski, 2013; Gocha et al., 2015; Hens & Belcastro, 2012; Hens & Godde, 2016; Merritt, 2015; Moraitis et al., 2014; Mulhern & Jones, 2005; Nagaoka & Hirata, 2008; Rissech et al., 2012; Storey, 2007; Wescott & Drew, 2015)

(Osborne et al., 2004)

Tests/applications: (Gocha et al., 2015)

(Passalacqua, 2009)

Tests/applications: (Colarusso, 2016)

Observations/other: (Ferrant et al., 2009; Igarashi, Uesu, Wakebe, & Kanazawa, 2005; A. Schmitt, 2005)

Combined Techniques

(Lovejoy, Meindl, Mensforth, et al., 1985)

Tests/applications: (Bedford et al., 1993; Fairgrieve & Oost, 1995; Galera et al., 1998)

(Acsadi & Nemeskeri, 1970)

Tests/applications: (Aiello & Molleson, 1993; Galera et al., 1998; Maat, 1987; Molleson, 1995; Perizonius, 1984; Saunders et al., 1992; Sjvold, 1975)

(A. Schmitt et al., 2002)

Tests/applications: (Martins et al., 2012)

(Boldsen et al., 2002)

Tests/applications: (Bullock et al., 2013; DeWitte, 2010; Godde & Hens, 2012; Hurst, 2010; Milner & Boldsen, 2012c; Storey, 2007; Wittschieber et al., 2014)

(Samworth & Gowland, 2007)

Tests & Applications: (Passalacqua, 2010)

Observations/other: (M. F. Anderson et al., 2010; Rougé-Maillart et al., 2009)

APPENDIX B: PHASE 1—TRANSITION CURVES FOR SELECTED FEATURES

In Phase 1, a sample of 1010 individuals from four well-documented skeletal collections were used to select a set of features that collectively show age-related variation throughout adulthood. This appendix contains transition plots for these features, which include both a regression curve fitted with the glm() function, a standard part of the program R, and a generalized additive model (GAM) generated with the gam() function in the R package mgcv (Team, 2008; S. N. Wood, 2004, 2006a). Figure 1.0 shows the key components of the transition plots constructed for the analysis of Phase 1 traits.

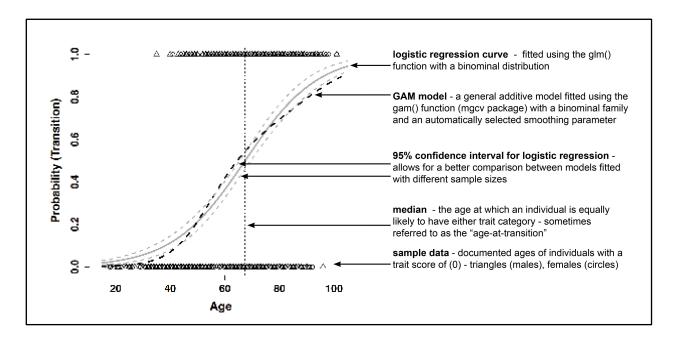


Figure B1.0. Key components of the transition plots used to evaluate Phase 1 traits.

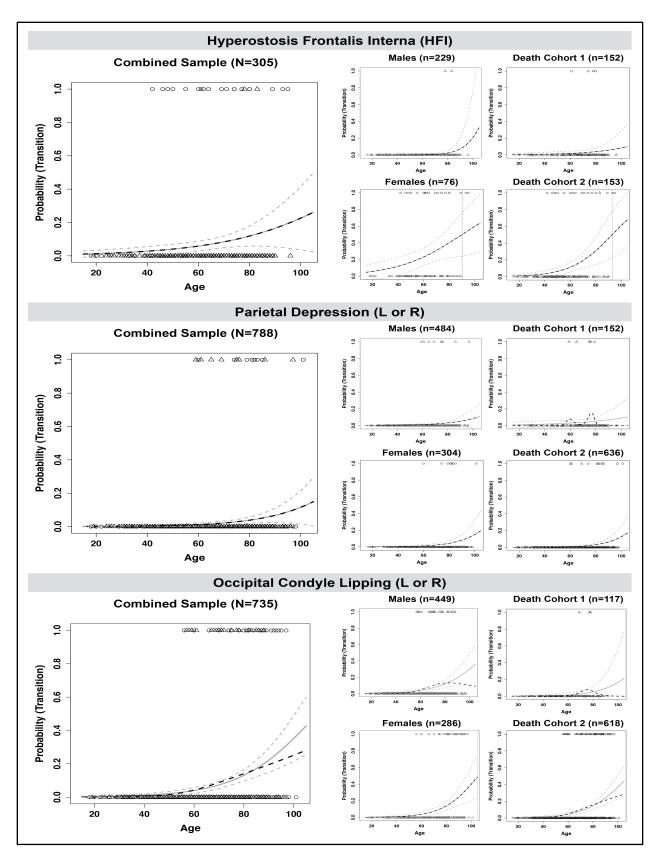


Figure B1.1. Transition plots for HFI, parietal depressions, and occipital condyle lipping.

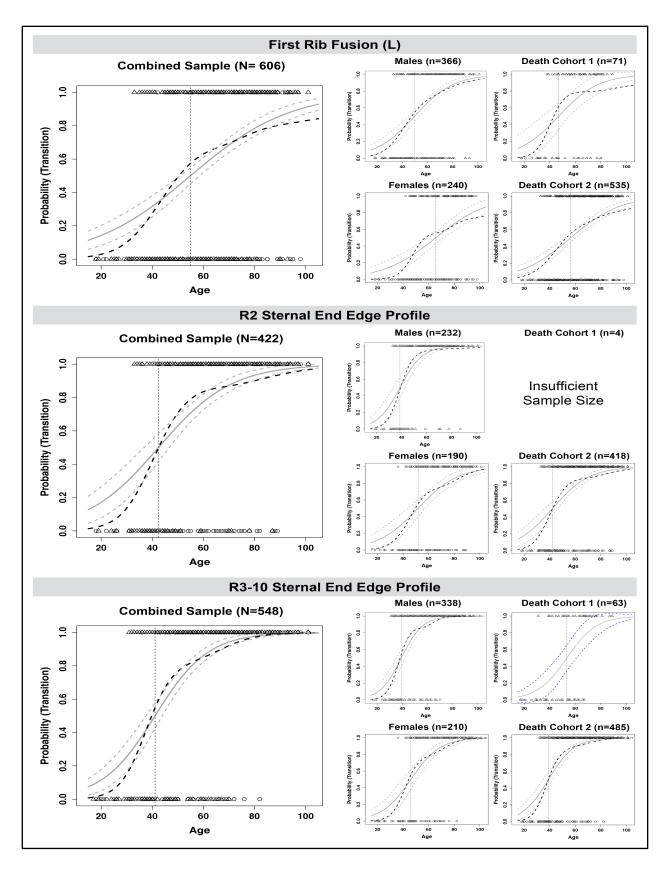


Figure B1.2. Transition plots for first rib fusion, R2 sternal end edge profile, and R3–10 sternal end edge profile.

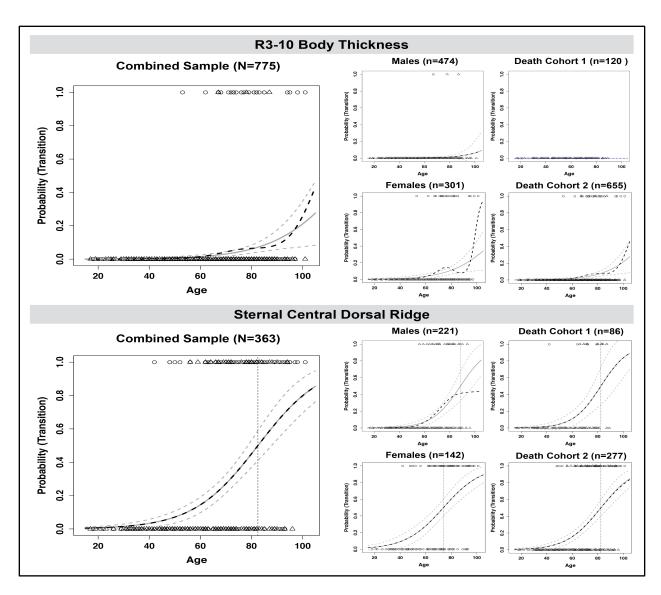


Figure B1.3. Transition plots for R3–10 body thickness ("shingle ribs") and sternal central dorsal ridge.

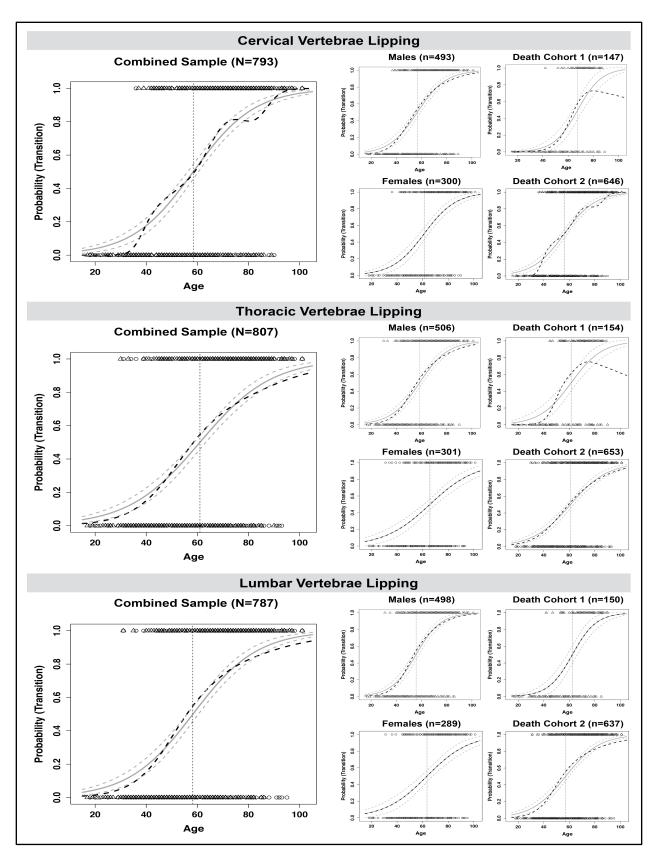


Figure B1.4. Transition plots for cervical, thoracic, and lumbar vertebral lipping.

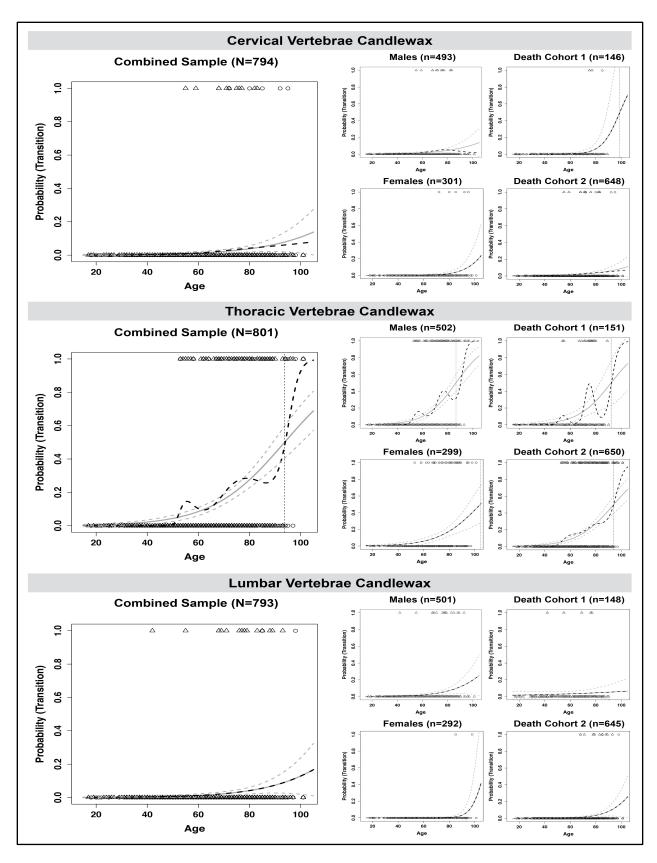


Figure B1.5. Transition plots for cervical, thoracic, and lumbar vertebral candlewax.

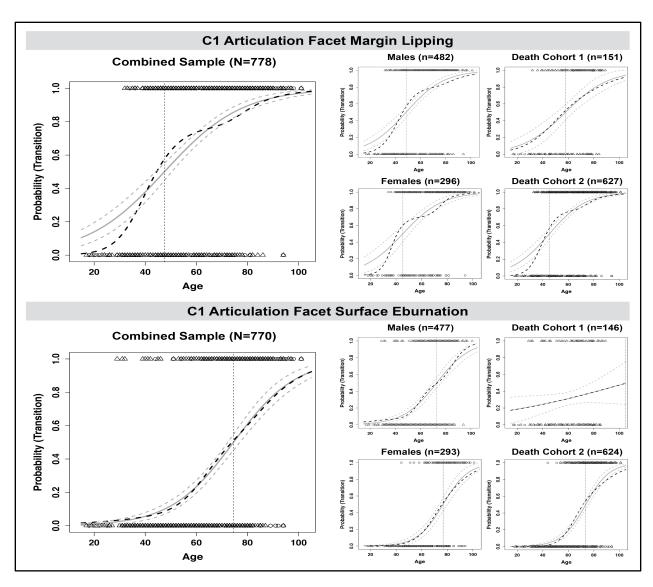


Figure B1.6. Transition plots for C1 articular facet margin lipping and C1 articulation facet surface eburnation.

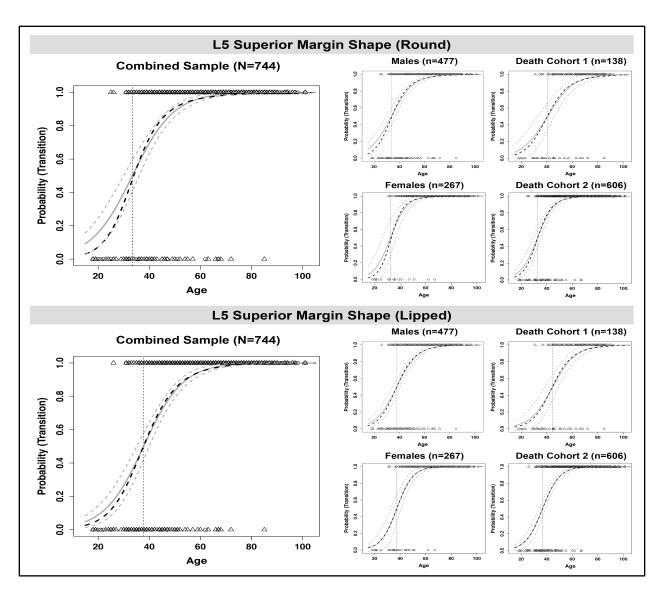


Figure B1.7. Transition plots for L5 superior margin shape.

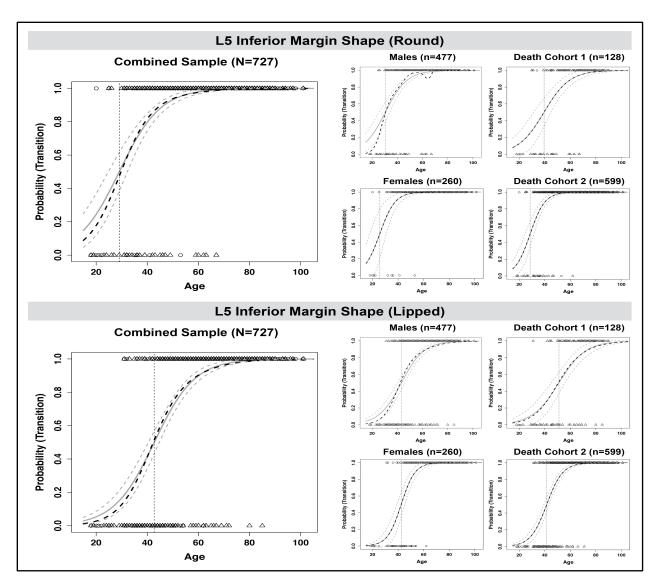


Figure B1.8. Transition plots for L5 inferior margin shape.

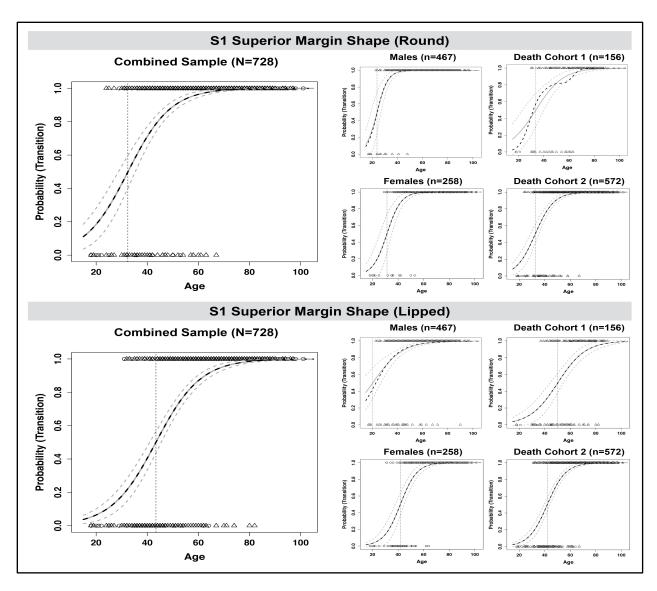


Figure B1.9. Transition plots for S1 superior margin shape.

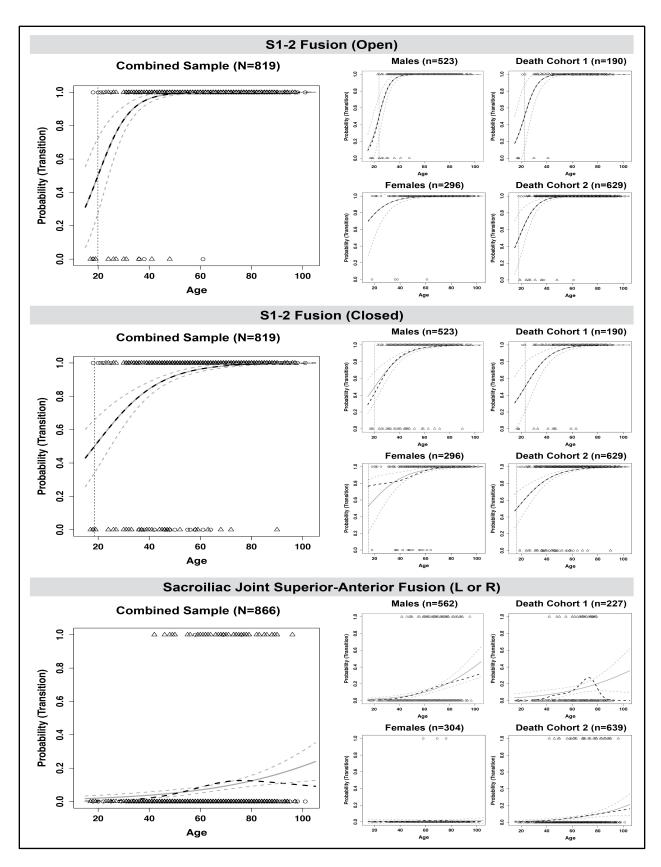


Figure B1.10. Transition plots for S1-2 fusion and superior anterior sacroiliac joint fusion.

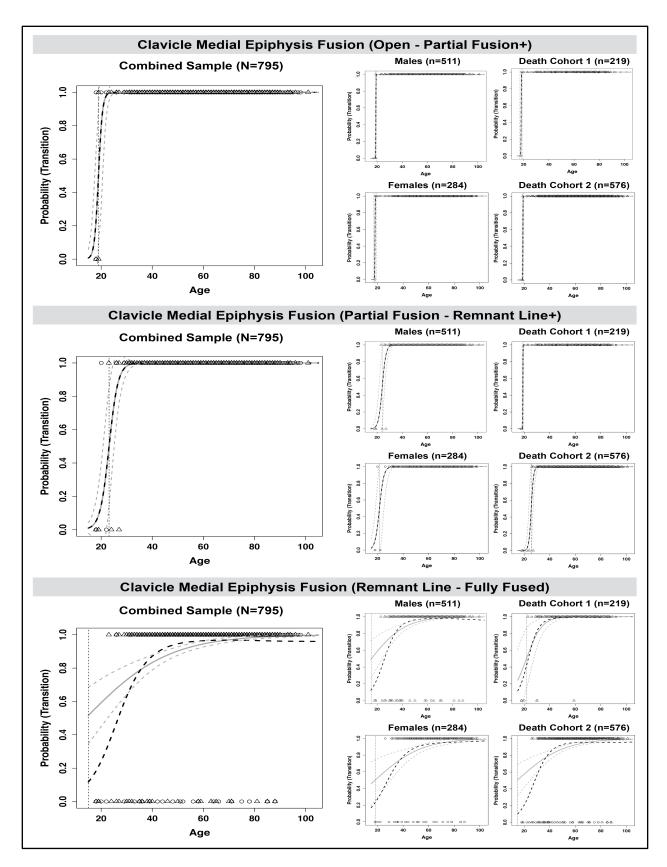


Figure B1.11. Transition plots for clavicle medial epiphysis fusion.

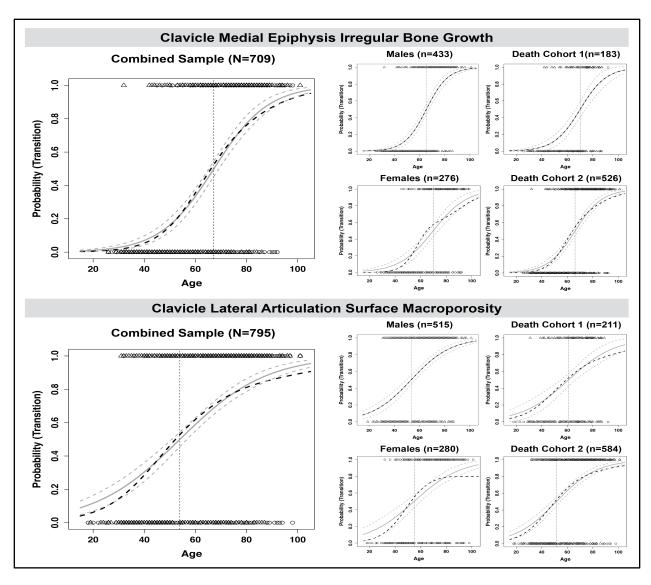


Figure B1.12. Transition plots for clavicle medial epiphysis irregular bone growth and clavicle lateral articulation surface macroporosity.

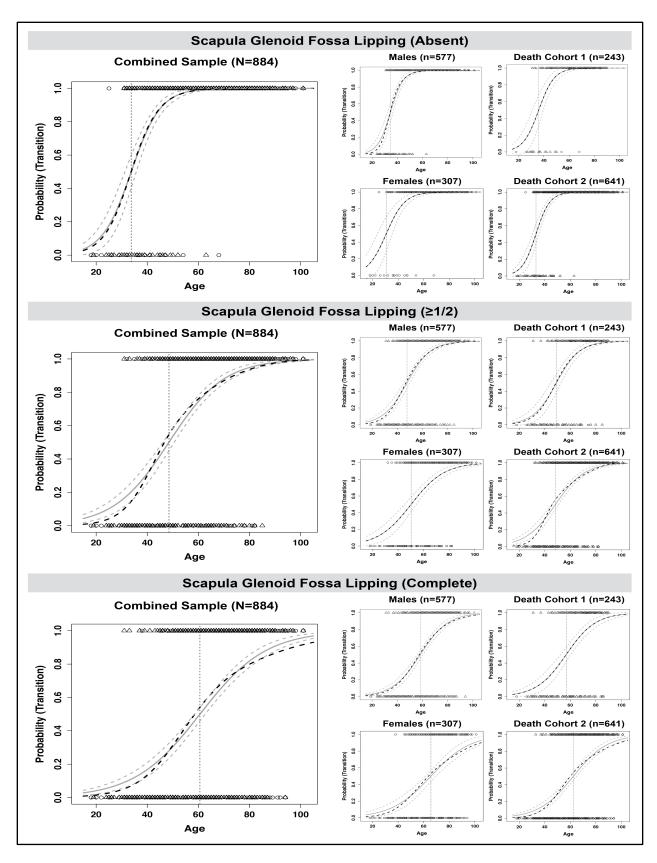


Figure B1.13. Transition plots for scapula glenoid fossa lipping.

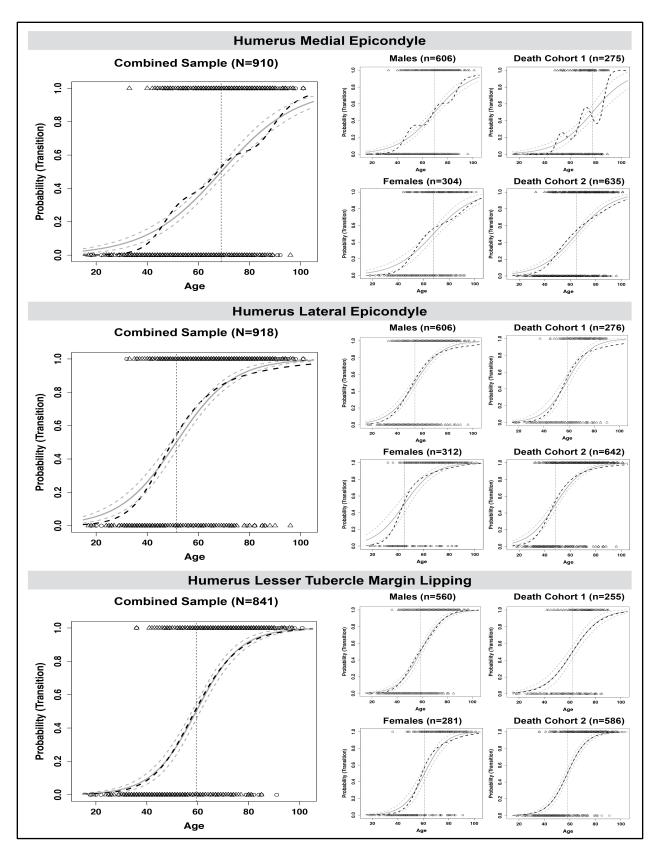


Figure B1.14. Transition plots for humerus medial epicondyle, humerus lateral epicondyle, and humerus lesser tubercle margin lipping.

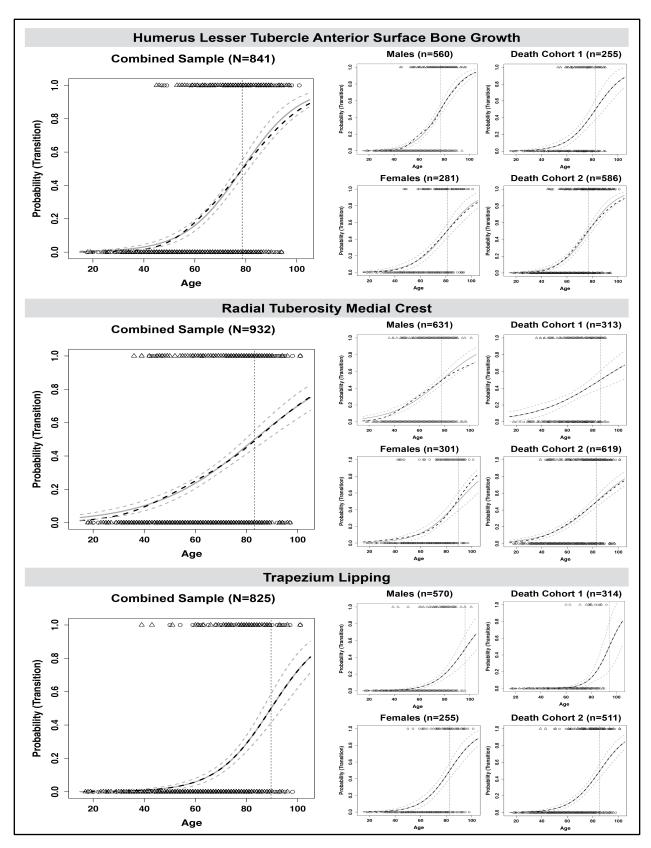


Figure B1.15. Transition plots for humerus lesser tubercle anterior surface bone growth, radius tuberosity medial crest, and trapezium lipping.

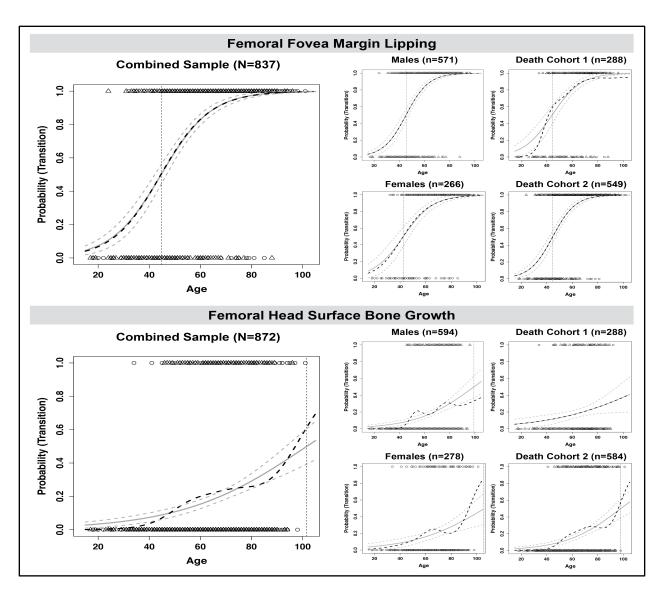


Figure B1.16. Transition plots for femoral fovea margin lipping and femoral head surface bone growth.

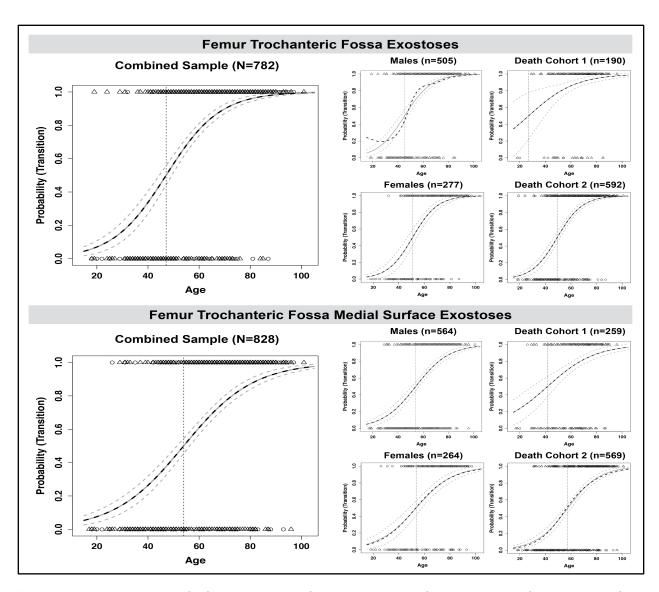


Figure B1.17. Transition plots for femur trochanteric fossa exostoses and femur trochanteric fossa medial surface exostoses.

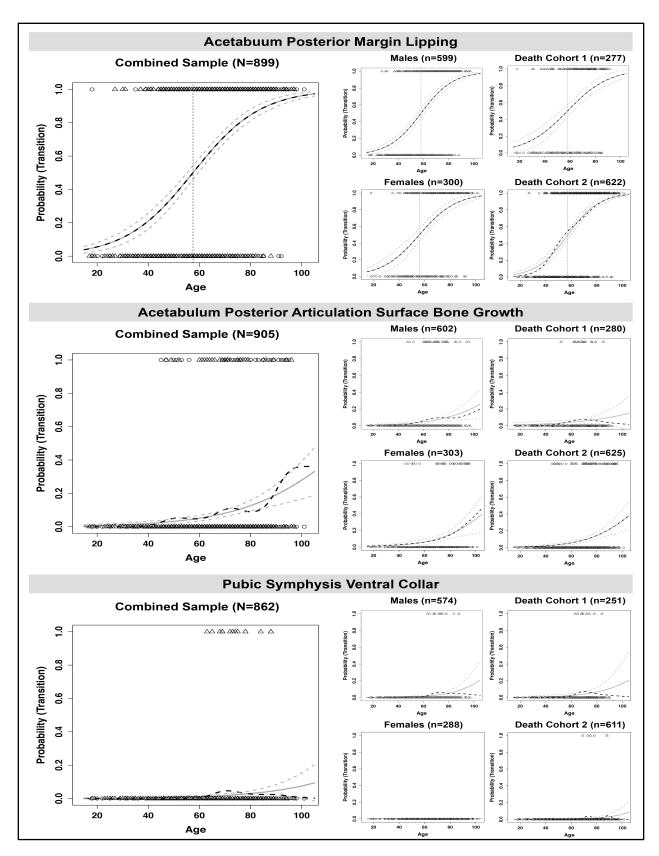


Figure B1.18. Transition plots for acetabulum posterior margin lipping, acetabulum posterior articulation surface bone growth, and pubic symphysis ventral collar.

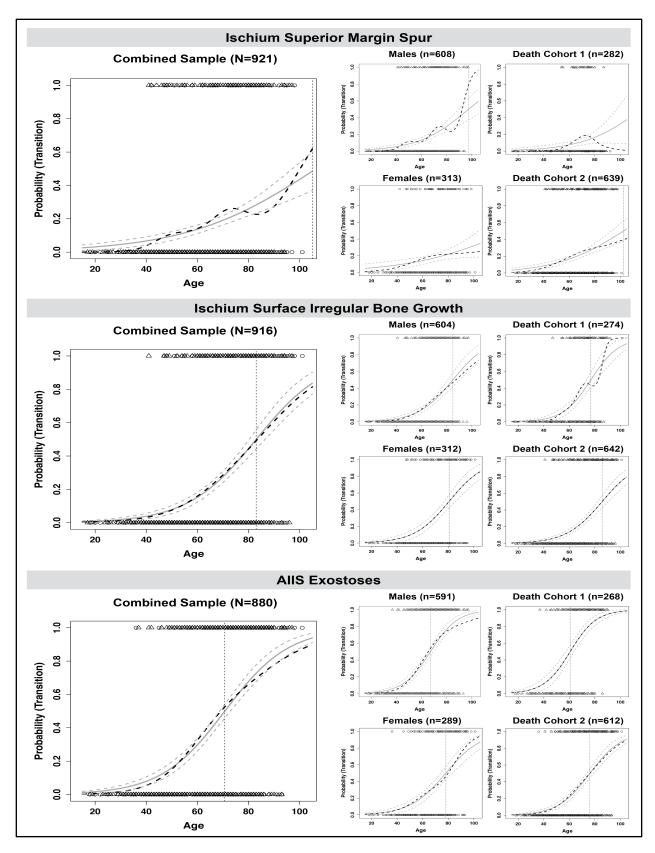


Figure B1.19. Transition plots for ischium superior margin spur, ischium surface irregular bone growth, and AIIS exostoses.

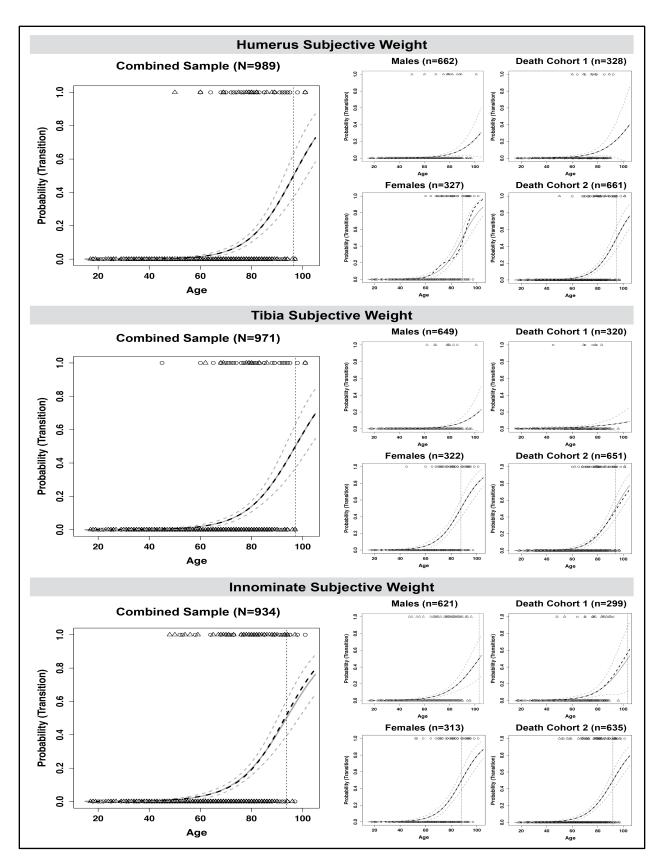


Figure B1.20. Transition plots for the subjective weight of the humerus, tibia, and innominate.

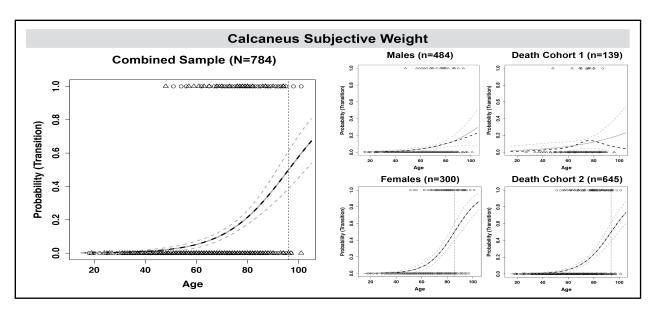


Figure B1.21. Transition plots for the subjective weight of the calcaneus.

APPENDIX C: TRANSITION ANALYSIS EQUATIONS

In logistic regression, the natural log of the odds of a trait being present (p) (C1.0)—is mathematically equivalent to a linear combination of the independent variables and an error term (ϵ) that is assumed to be independent of age (C1.1). The basic mathematical steps to calculate probability from the fitted logistic model are shown in steps (C1.2–C1.8) and described briefly in Table C1.0⁴⁰.

$$logit(p) = ln(odds) = ln\left(\frac{p}{1-p}\right) = ln\left(\frac{probability\ of\ trait\ present)}{probability\ of\ trait\ absent}\right)$$
 (C1.0)

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \varepsilon$$
 (C1.1)

$$e^{\ln\left(\frac{p}{1-p}\right)} = e^{\beta_0 + \beta_1 x} \tag{C1.2}$$

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x} \tag{C1.3}$$

$$p = e^{\beta_0 + \beta_1 x} (1 - p) \tag{C1.4}$$

$$p = e^{\beta_0 + \beta_1 x} - p \left(e^{\beta_0 + \beta_1 x} \right) \tag{C1.5}$$

$$p + p(e^{\beta_0 + \beta_1 x}) = e^{\beta_0 + \beta_1 x}$$
 (C1.6)

$$p(e^{\beta_0 + \beta_1 x} + 1) = e^{\beta_0 + \beta_1 x}$$
 (C1.7)

$$p = \frac{e^{\beta_0 + \beta_1 x}}{(e^{\beta_0 + \beta_1 x} + 1)}$$
 (C1.8)

Table C1.0. Steps to calculate probability from the logistic model.

Starting Equation	Action	Resulting Equation
$ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$	Take the exponential of each side.	$\frac{p}{(1-p)} = e^{\beta_0 + \beta_1 x}$
$\left(\frac{p}{1-p}\right) = e^{\beta_0 + \beta_1 x}$	Multiple each side by (1-p)	$p = e^{\beta_0 + \beta_1 x} \left(1 - p \right)$
$p = e^{\beta_0 + \beta_1 x} \left(1 - p \right)$	Distribute.	$p = e^{\beta_0 + \beta_1 x} - p (e^{\beta_0 + \beta_1 x})$
$p = e^{\beta_0 + \beta_1 x} - p \left(e^{\beta_0 + \beta_1 x} \right)$	Add $p(e^{\beta_0 + \beta_1 x})$ to both sides.	$p + p\left(e^{\beta_0 + \beta_1 x}\right) = e^{\beta_0 + \beta_1 x}$
$p + p\left(e^{\beta_0 + \beta_1 x}\right) = e^{\beta_0 + \beta_1 x}$	Take out p.	$p(e^{\beta_0 + \beta_1 x} + 1) = e^{\beta_0 + \beta_1 x}$
$p(e^{\beta_0 + \beta_1 x} + 1) = e^{\beta_0 + \beta_1 x}$	Divide by $(e^{\beta_0 + \beta_1 x} + 1)$.	$p = \frac{e^{\beta_0 + \beta_1 x}}{(e^{\beta_0 + \beta_1 x} + 1)}$

⁴⁰ For clarity, the error term is not shown in these steps.

-

As shown in the above section, the probability of a trait being present at a specific age can be recovered from the logistic model by using the logit link function (C1.19), which we will represent in the following section as (F). The probability of a trait being preset at age (x) using the logit link function is shown in equation C1.10. Because probabilities must sum to one, the probability of a trait being absent at the same age is equal to one minus the probability of it being present (C1.11).

logit link function =
$$F = \frac{e(\cdot)}{1+e(\cdot)}$$
 (C1.9)

$$Pr(y=1 \mid x) = \frac{e^{\beta_0 + \beta_1 x}}{(e^{\beta_0 + \beta_1 x} + 1)} = F(\beta_0 + \beta_1 x)$$
 (C1.10)

$$Pr(y = 0 \mid x) = 1 - \frac{e^{\beta_0 + \beta_1 x}}{(e^{\beta_0 + \beta_1 x} + 1)} = 1 - F(\beta_0 + \beta_1 x)$$
 (C1.11)

If the features are independent, which in the case of adult age estimation is always a questionable assumption, the combined likelihood function for the presence or absence of multiple traits is the product of their individual likelihood functions (C1.12). For simplicity, this equation can be rewritten as equation C1.13.

$$L = \prod_{i=1}^{N} \left[F(\beta_0 + \beta_1 x_i) \right]^{y_i} \left[1 - F(\beta_0 + \beta_1 x_i) \right]^{1-y_i}$$
 (C1.12)

$$L = \prod_{i=1}^{N} [F_i]^{y_i} [1 - F_i]^{1 - y_i}$$
 (C1.13)

To make this equation easier to deal with from a mathematical and practical standpoint, we take the natural log of both sides (C1.14). Using the properties of logarithms, this equation can be rearranged into a more useful form (C1.15–C1.19). Equation C1.19 is ultimately what is used to produce the maximum likelihood age estimates for the validation study in Phase 3.

$$Ln L = Ln [(F_1)^{y_1} (1 - F_1)^{1 - y_1} + (F_2)^{y_2} (1 - F_2)^{1 - y_2} + \dots + (F_n)^{y_n} (1 - F_n)^{1 - y_n}]$$
 (C1.14)

$$Ln L = Ln [(F_1)^{y_1} (1 - F_1)^{1-y_1}] + Ln [(F_2)^{y_2} (1 - F_2)^{1-y_2}] + \dots + Ln [(F_n)^{y_n} (1 - F_n)^{1-y_n})]$$
(C1.15)

$$Ln L = Ln [(F_1)^{y_1}] + Ln[(1 - F_1)^{1-y_1}] + Ln[(F_2)^{y_2}] + Ln[(1 - F_2)^{1-y_2}] + \dots + Ln [(F_n)^{y_n}] + Ln[(1 - F_n)^{1-y_n}]$$
(C1.16)

$$Ln L = y_1 Ln (F_1) + (1 - y_1) Ln (1 - F_1) + y_2 Ln (F_2) + (1 - y_2) Ln (1 - F_2) + \dots + y_n Ln (F_n) + (1 - y_n) Ln (1 - F_n)$$
(C1.17)

$$Ln L = \sum_{i=1}^{n} y_i \ln(F_i) + (1 - y_i) \ln(1 - F_i)$$
 (C1.18)

$$Ln L = \sum_{i=1}^{n} y_i \ln(Pr(y_i = 1 \mid x_i) + (1 - y_i) \ln(Pr(y_i = 0 \mid x_i))$$
(C1.19)

Revised Transition Analysis

Trait Manual Ver 3.0



George R. Milner Sara M. Getz



Stephen D. Ousley Svenja Weise



Jesper L. Boldsen Peter Tarp

Figure D1.0. NIJ trait manual cover page.

Parietal depression [bilateral]

Location: the parietal bone between the superior temporal line and the sagittal suture (PD Fig. 1)

Scores: 0. absent 1. present

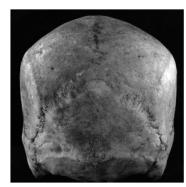
Absent: The parietal bone is rounded or flattened, but not depressed (PD Fig. 2).

Present: A portion of the parietal bone is slightly depressed (concave) relative to its normal contour. A depression is most often a rounded or oval area located on the middle to posterior portion of the parietal bone. Less typically, the depression can be expressed as a long groove, two or more centimeters wide, just superior to the temporal line.

Notes:

- The contour of the parietal bones is most easily assessed in a posterior view. However, depressions can often be seen in a superior view, particularly when the bone where the depression is located is particularly thin (e.g., PD5, PD8)
- Depressions can be subtle, asymmetric, and may occur unilaterally.
- Robust parietal bosses, well-developed temporal lines, sagittal keeling, or a combination of these features can cause parietal bones to appear depressed. Take care to differentiate these features from parietal bones that are depressed relative to the normal contour of the parietal bone.
- Well-healed cranial trauma (unilateral or bilateral) is distinctively different than parietal thinning and should not be scored.

Examples:



PD-1. Absent (posterior view) The parietal bone is rounded, but not depressed. (CM160-52, 46F)





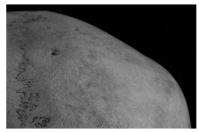
PD Fig 1: Parietal depression scoring location







PD Fig 2: Parietal bone contour variants: a) round, b) flattened, c) depressed.



PD-2. Absent (posterior view) The parietal bone is flattened, but not depressed. (UP1046-6983, 86F)



PD-3. Present (posterior view) Subtle bilateral depressions. (UT13-05, 75F)



PD-4. **Present** (a, posterior view) - bilateral, depressions.

(b, superior view) - left depression is an elongated groove just superior to the temporal line. (UT20-01, 83F)

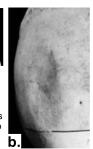


Figure D1.1. Parietal depression description and examples (page 1 of 2).

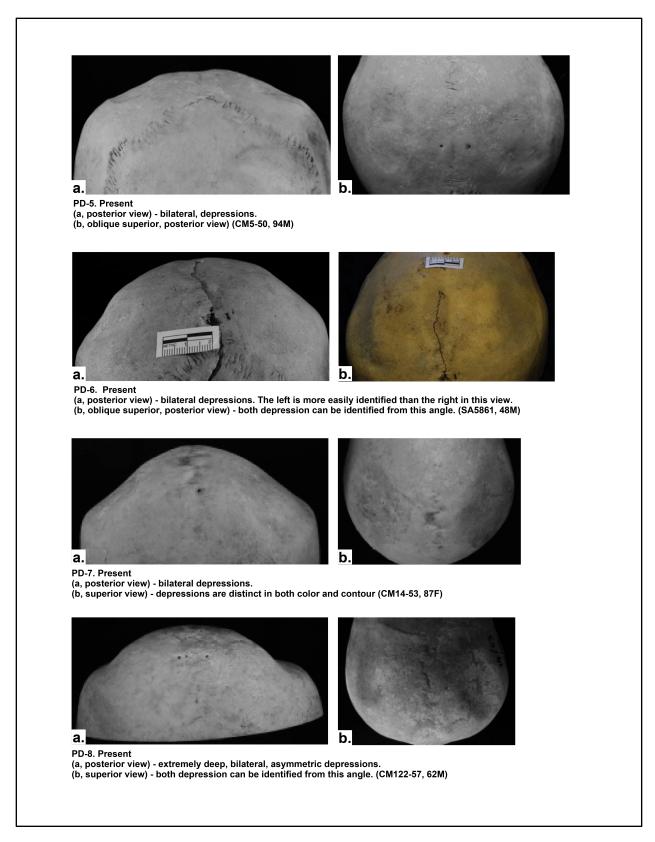


Figure D1.2. Parietal depression description and examples (page 2 of 2)

C1 lipping (1st cervical vertebra dens articulation surface margin lipping)

Location: the edges of the articulation surface for the dens (odontoid process) on the posterior surface of the anterior arch of the first cervical vertebra (atlas) (C1L Fig. 1)

Scores: 0. absent (<50%) 1. present (≥50%)

Present (≥50%): Irregular bony growth extends superiorly, inferiorly, or laterally ≥1.5 mm from the joint margin, bordering ≥50% of the articulation surface.

Absent (<50%): It is common to have a rounded to sharp, distinct, but narrow and regular, bony elevation surrounding part or all of the facet. This raised rim can sometimes take on a somewhat irregular appearance, but does not satisfy the length or extension criteria for lipping.

Notes:

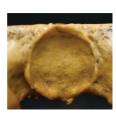
- The bone growth around the articulation facet resembles the lipping seen at other synovial joint margins.
- The present category is typically easily identifiable as soon as you pick up the bone.
 If close inspection is required, then the trait is usually scored as absent.



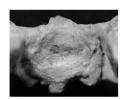
C1L-1: **Absent**The articulation facet does not have distinct margins.
No lipping is present.
(CM340-56, 15F)



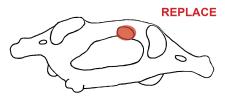
C1L-2: **Absent** Small raised rim inferiorly. No lipping is present. (CM64-51, 76M)



C1L-3: **Absent**Narrow raised rim that does not meet the criteria for lipping to be present. (UT02-87, 75M)



C1L-4: **Present**At least 1.5mm of lipping extends outward around ≥50% of the articular facet. (CM112-53, 51M)



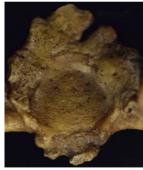
C1L Fig. 1: Location of dens articular facet.



C1L-5: **Present** (UT09-06, 35M)



C1L-6: **Present**More than 1.5mm of lipping extends outward from around more than half of the articular facet.
(CM135-46, 90M)



C1L-7: **Present** (UT78-08, 92F)

Figure D1.3. C1 articular facet surface margin lipping description and examples.

C1 eburnation (C1 dens articulation surface eburnation)

Location: the articulation surface for the dens on the anterior arch of first cervical vertebra (C1E Fig. 1)

Scores: 0. absent 1. present

Present: Eburnation is characterized an area of shiny, smooth, dense-looking bone on either the joint surface or any lipping surrounding it. A single spot of eburnation is sufficient to be scored as present.

Absent: No eburnation exists on either the articulation facet or the lipping surrounding it.

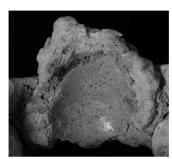
Notes:

- C1 eburnation is the same as that often seen on the articulation surfaces of other synovial joints.
- Any bony growth around the articulation surface should be treated as an extended joint surface. If eburnation is found only on this lipping, it should still be scored as present.
- If the bone is covered in preservative, dirt, or grease, this feature is often not observable.
- It is possible to have eburnation without being scored as ≥50% lipped (see C1 lipping).

Examples:

Photo Not Yet Available

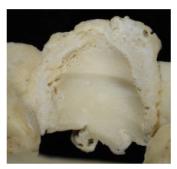
C1E-1: **Present** Small area of eburnation present.



C1E-2: **Present** Lipped with extensive eburnation. UP1384-5570, 83M)



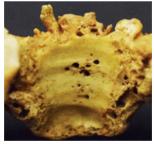
C1E Fig. 1: Location of dens articular facet.



C1E-3: **Present** Lipping with extensive eburnation. (CM53-50, 44M)



C1E-4: **Present** Lipping with extensive eburnation. (CM135-46, 90M)



C1E-5: **Present**Original articular surface and lipping eburnated.
(UT88-09, 82M)

Figure D1.4. C1 articular facet surface eburnation description and examples.

L5 margin shape (superior & inferior)

Location: the anterior 20 mm (10 mm either side of the midline) of the superior and inferior margins of the fifth lumbar vertebra (L5MS Fig. 1).

Scores: 0. round 1. sharp 2. lipped

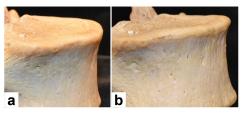
Round: The ventral margin is smooth and rounded. The margin transitions without interruption from the superior or inferior surface to the anterior surface.

Sharp: A reasonably sharp edge (angle) is present between the articular and anterior surfaces. The edge feels sharp when rubbed against the sensitive skin on the palmar surface of the finger joints.

Lipped: The margin has any amount of bony lipping. Lipping is typically occurs on an otherwise angular margin. If lipping has resulted in fusion to another vertebral element, it should be scored as lipped.

Notes:

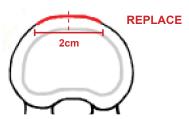
- It is sometimes easier to score the inferior margin by turning the vertebra over to view the margin in an inferior to superior direction.
- If the epiphyseal ring is unfused, the margin should be scored as *round*.



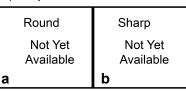
L5MS-1: **Round**Lateral (a) and anterior-oblique (b) views of the same superior margin. The superior surface merges with the anterior face in a smooth, continous arc. (UT20-90, 29M).



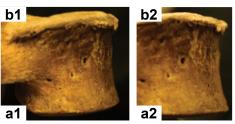
L5MS-2: **Round**Both the superior and inferior margins are round.
(UP1754-6058, 30M)



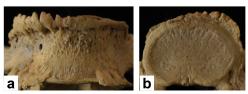
L5MS Fig. 1: Anterior 20 mm (2cm) of the vertebral margin (superior margin shown). Superior and inferior margins are scored separately.



L5MS Fig. 2: Profiles of (a) round (b) sharp margins. Lipping typically occurs on a sharp (angular) margin.



L5MS-3: **Sharp (a) & lipped (b)**Lateral (1) and anterior-oblique (2) views of a sharp inferior margin (a) and a lipped superior margin (b). Inferiorly, the anterior surface of the body meets the inferior articular surface at an angle, not a smooth curve (compare to L5MS-1).



L5MS-4: **Lipped**Anterior (a) and superior (b) views of a lipped superior margin. (UP1525-5022, 41M)



L5MS-5: **Lipped**Superior view of lipped superior margin. (UT32-09, 63M)

Figure D1.5. L5 margin shape description and examples.

S1 margin shape (1st sacral vertebra superior anterior border)

Location: middle 20 mm of the anterior (ventral) superior edge of the first sacral body (S1)

Scores: 0. round 1. sharp 2. lipped

Round: The (ventral) margin is smooth and rounded. The margin transitions without interruption from the superior to anterior surface.

Sharp: A reasonably sharp edge (an angle) is present between the superior and anterior surfaces.

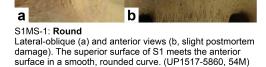
Lipped: The margin has any amount of bony lipping Lipping is typically only found on an otherwise angular margin.

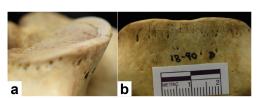
Notes:

Examples:

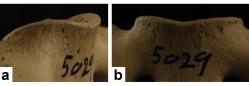
- The difference between rounded and sharp must be adjusted for size. It is common for particularly small individuals to have a margin that feels sharper than normal, but that is still round.
- The distinction between the round and sharp categories can be felt as well as seen – the former is best done with the sensitive part of the palmar surface of a finger joint.
- The bone should be scored as lipped if fusion to the last lumbar vertebra has occurred as the result of bony lipping.

CONTRACTOR OF THE PARTY OF THE





S1MS-2: **Round**RoundLateral-oblique (a) and anterior views (b). Although less round than (S1MS-1), the superior articular surface of S1 merges smoothly with the anterior face of the sacrum. No distinctly sharp angle is present. (UT18-90, 27M)



S1MS-3: **Round**Lateral-oblique (a) and anterior views (b). Although less round than (S1MS-1), the superior articular surface of S1 merges smoothly with the anterior face of the sacrum. No distinct sharp edge (angle) is present. (UP1152-5029, 16F)





S1MS-4: Round (a) & sharp (b)
Oblique views of round (a) and sharp (b) margins. In sacrum
(b), the anterior margin where the superior and anterior
surfaces meet is a sharp angle. This distinction can be seen
and felt. [a) CM110-57,19F; b) CM338-56,38M]



S1MS-5: **Lipped.**Superior-oblique view of a slightly lipped margin. Lipping occurs on an otherwise angular margin; any lipping is sufficient for a lipped score. (UT06-87, 69M)



S1MS-6: **Lipped** Anterior view of lipped margin. (CM1440-6486, 84F)

Figure D1.6. S1 margin shape description and examples.

S1-2 fusion (1-2 sacral body fusion)

Location: the anterior surface of the sacrum where the irst two sacral bodies meet

Scores: 0. gap ≥10mm 1. gap <10mm 2. closed

Gap ≥10mm: A gap more than 10 mm in length is open without interruption.

Gap <10mm: A gap less than 10mm in length is present.

Closed: There is no evidence of a gap between the anterior sacral bodies.

Notes:

- A gap is considered present when there is an opening that extends into the space between the two sacral bodies.
- Many specimens where vertebral shifts occur cannot be recorded.

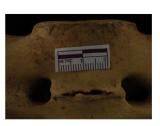
Examples:



S12F-1: **Gap ≥10mm** (UT14-04, 19M)



S12F-2: **Gap ≥10mm** UT21-92, 25M)



S12F-3 **Gap ≥10mm** (UP1008-5885, 32F)



S12F-4: **Gap ≥10mm**An opening just over 10mm in length is present. (UT18-90, 27M)



S12F-5: **Gap <10mm** (UP1712-6186, 39M)



S12F-6: **Closed** (UP1603-2908, 49F)

Photo Not Yet Available

S12F-7: **Closed**Gap present with fused bottom.

Figure D1.7. S1-2 fusion description and examples.

Vertebral lipping

Location: the anterior portions of the superior and inferior edges of cervical vertebrae three through seven, and in all thoracic and lumbar vertebrae

Scores:

 cervical: 0. absent (<3 mm)</td>
 1. present (≥3 mm)

 thoracic: 0. absent (<5 mm)</td>
 1. present (≥5 mm)

 lumbar: 0. absent (<5 mm)</td>
 1. present (≥5 mm)

Cervical, thoracic, and lumbar vertebrae are scored separately. The number of scorable edges in each group is recorded.

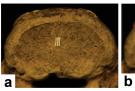
Present: Lipping is present if it extends anteriorly, cranially, or caudally from the margin ≥3 mm for cervical or ≥5 mm for thoracic and lumbar vertebrae. Lipping is measured from the original vertebral margin to the furthest extent of the osteophyte (outward, upward, or downward from the vertebral body), not the distance along the margin.

Notes:

- Vertebrae fused through lipping or candlewax should be scored as lipped (e.g., VL-7).
- Congenitally fused vertebrae, particularly common in the cervical spine, should not be scored.
- It is not unusual for vertebrae to appear highly irregular and porous, but have an insufficient amount of lipping to be scored as present. This is particularly common for cervical vertebrae.



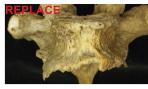
VL-1: **Absent**The lipping on this lumbar vertebra extends outward from the original edge by <5 mm. (UT08-89, 66M)



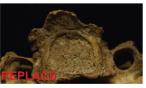
VL-2: **Absent**Lipping extends outward from the original edge by <5 mm.
The same vertebra is shown in (b) with the original margin (dotted line) and lipping (solid line). (UT33-10, 40M)



VL-3: Lipped (≥5 mm) Superior view of lumbar vertebra. (UP1647-5849, 83M)



VL-4: Lipped (≥5 mm)
Anterior view of a thoracic vertebra where lipping extends inferiorly for ≥5 mm. (UT53-04, 83M)



VL-5: **Lipped (≥3 mm)** Inferior view of a cervical vertebra with lipping that extends ≥3 mm. (UT78-08, 92F)



VL-6: **Lipped (≥5 mm)**Thoracic lipping that is also candlewax. (UT94-06, 93F)



VL-7: Lipped (≥3 mm)
Fusion of two cervical vertebrae through candlewax. Fusionin this manner is sufficient to be scored as lipped, however, the inferior of the fused vertebrea also has sufficient lipping to independently be scored as lipped. (UT78-08, 92F)

Figure D1.8. Cervical, thoracic, and lumbar vertebral lipping description and examples.

Vertebral candlewax

Location: superior edges, inferior edges and anterior surfaces of the vertebral bodies of cervical three through seven, and all thoracic and lumbar vertebrae

Scores: 0. absent 1. present

Cervical, thoracic, and lumbar vertebrae are scored separately. The number of scorable edges is recorded.

Present: Candlewax is a distinctive form of bony growth typically associated with advanced osteoarthritis and Diffuse Idiopathic Skeletal Hyperostosis (DISH). The hard and smooth (sclerotic) bone looks like hardened candlewax that has flowed down the vertebral bodies and across the joints between vertebrae. Candlewax expands at the joints of vertebrae, creating swellings in the bony flow. If candlewax is visible on one or more of the vertebral edges, then the trait is scored as present.

Notes:

- Fusion of vertebral elements is not necessary for candlewax to be present (e.g., CW-2, CW-3)
- As compared to osteophytes, which originate from the margins of the vertebrae, the smooth, shiny sclerotic bone, characteristic of candlewax is typically present on the anterior surfaces of the vertebrae, as well as flowing over and across the joint margins.

Examples:

Photo Not Yet Available CW-1: **Absent** Lipping that does not meet the qualifications to be candlewax.



CW-2: Present Minor, unfused candlewax. The candlewax originates on the body of the vertebrae, spans the joint space, and has a different texture than the surrounding bone. (UP1286-5191, 60F)





CW-3: **Present**Anterior (a) and lateral-oblique (b) views of unfused candlewax present only on the left inferior edges. The candlewax originates on the body of the vertebra, spans the joint space, and has a sclerotic appearance. (UP1371-5221, 57M)



CW-4: Present
Candlewax of adjacent
vertebrae spans the joint
space and forms touching
facets, but is not fused. It is
different in both color and
texture than the surrounding
bone. (UT94-06, 93F)





CW-5: **Present**Lateral (a) and anterior (b) views of candlewax-fused vertebrae. (UT58-09, 81M)

Figure D1.9. Cervical, thoracic, and lumbar vertebral candlewax description and examples.

Sternum central dorsal ridge

Location: the dorsal (posterior) surface of the completely fused adult sternal body, excluding the manubrium and xiphoid process

Scores: 0. absent 1. present

Present: A single, central ridge runs superiorly to inferiorly along the midline of the dorsal sternal surface. It must be 15 mm long and 1.5 mm wide at least at one point.

The ridge is an addition of bone on the sternal surface with sharp borders. It can be seen and felt (compare SDR-2 and SDR-6).

Notes:

- A line of small, narrow, linearly oriented spicules that does not form a continuous ridge is insufficient to be scored as present.
- A central ridge can be present in isolation or with similar lines or ridges that occur on the lateral borders of the sternal body. These laterally located lines can exist unilaterally, and should not be confused with the central dorsal ridge along the midline.
- The feature cannot be scored if there has been thoracic surgery that involved cutting the sternum superiorly to inferiorly.



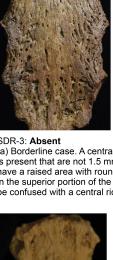
SDR-1: **Absent**No central ridge is present.
Poorly expressed lateral
ridges are present inferiorly,
but only ridges located in
the center of the dorsal
sternum are scored.
(UT02-91, 81M)



SDR-2: Absent
This is a slightly rounded, raised mound along the midline, but no substantial addition of bone. Compare to SDR-10. (UP1014-7049, 93M)



SDR-3: **Absent**(a) Borderline case. A central line of narrow, linear spicules is present that are not 1.5 mm wide. (b) It is common to have a raised area with rounded borders 10 to 20 mm long in the superior portion of the sternal body (b) that should not be confused with a central ridge. (UT25-06, 44F)



SDR-4: Absent Borderline case. The central dorsal ridge is present but is not at least 15mm in long. Lateral ridges are also present, but are not scored as part of this trait. (UP1366-6474, 85F)



SDR-5: **Absent**A single lateral ridge is present on the left side, but no central ossification is present. (UT78-08, 92F)

Figure D1.10. Sternum central dorsal ridge description and examples (page 1 of 2).

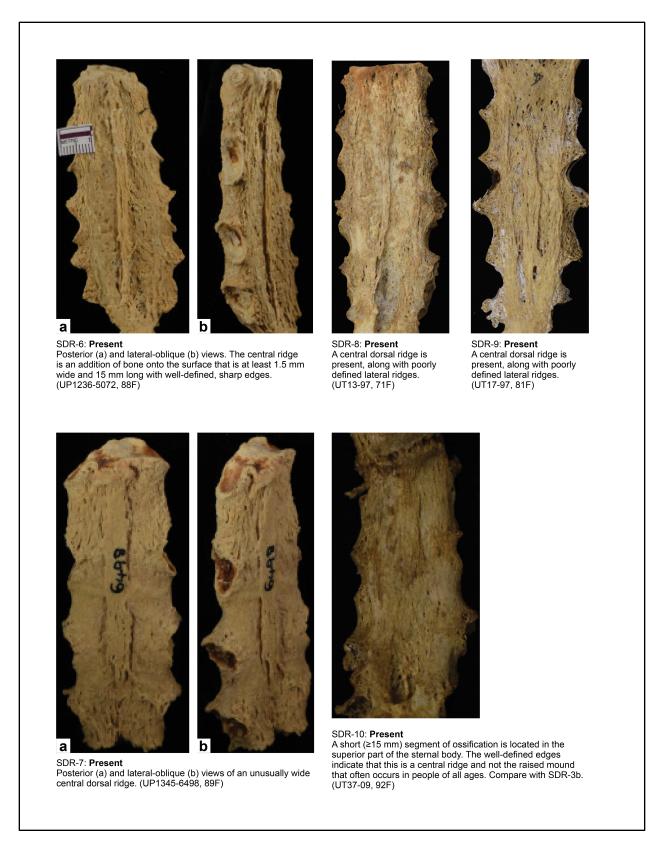


Figure D1.11. Sternum central dorsal ridge (page 2 of 2) description and examples (page 2 of 2).

R2 & R3-10 sternal end rim profiles

Location: the sternal end of rib two (R2) and ribs three through ten (R3-10), specifically the oval margin of each rib end where bone meets costal cartilage

Scores: 0. regular 1. irregular

Regular: The end of the rib is straight or scalloped, but not lipped.

Irregular: The margins of the sternal end have thinned and at least a small amount of bone growth results in an irregular appearance. In many cases, additional irregular ossifications project beyond the original borders of the rim resulting in a ragged or claw-like appearance.

Notes:

 Occasionally tube-like ossifications encircle the entire sternal end of a rib, but do not form jagged (or claw-like) protrusions. These ossifications of cartilage are also classified as irregular.

Examples:



RP-1: **Regular** Regular rim with well-defined scallops. (UP1008-5885, 32F)



RP-2: **Regular** The edges are thick and rounded. (UT21-92, 25M)



RP-3: **Regular**The edges are rounded and the done is dense with a slightly scalloped appearance.
(UT20-90, 29M)



RP-5: Irregular (UT06-87, 69M)

Photo Not Yet

RP-4: Regular
Although a central ossification (cone) is present, the rim can still be considered regular.

Available



RP-6: Irregular (UT02-91, 81M)



RP-7: Irregular (UT06-87, 69M)



RP-9: **Irregular** (CM121-51, 80M)



RP-11: **Irregular** Rib with long claw-like ossifications extending several centimeters. (CM121-51, 80M)



RP-8: **Irregular** Preserved cartilage. (UT06-87, 69M)



RP-10: Irregular Tube-like ossification. Arrow indicates the approximate location of original rib end. (CM47-51, 81M)



RP-12 **Irregular**Ossification that takes the form of a cylinder or tube.
Arrow indicates the original end of the rib.
(UP1252-6008, 48M)

Figure D1.12. R2 and R3-10 sternal end rim profiles description and examples.

R3-10 shingle ribs (rib body thickness)

Location: the ventral third oribs three through ten (R3-10)

Scores: 0. normal 1. shingle

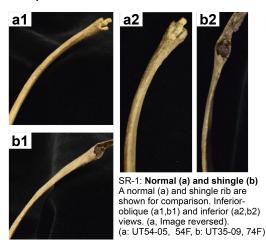
The number of scorable rib bodies (i.e., those with at least the anterior one-third intact) should be recorded.

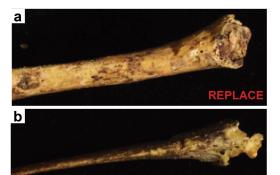
Shingle: Shingle ribs are quite narrow in cross section, and taper to a sharp margin both superiorly and inferiorly. The cortices are thin, especially noticeable in broken bones, and trabeculae are sparse. In cross section, the exterior and interior sides of the ribs parallel to one another, instead of the usual oval shape found in most adults. If present, the flattening generally occurs in the most distal (costal) thirds of the ribs, but can extend to as much as the distal two-thirds of the rib body.

Normal: The bone has a thick and robust appearance. The cross sections of ribs form an elongated oval with a rounded superior border and, typically, a sharp inferior border adjacent to the costal groove.

Notes:

- The overwhelming majority of individuals have normal ribs. Shingle ribs are only present in clear-cut cases of thinning.
- Shingle ribs are defined by their overall crosssectional shape, not cortical thickness. However, all shingle ribs have extremely thin cortices.





SR-2: **Normal (a) and shingle (b)**Comparison of a normal rib and a shingle rib. The sides of the shingle rib are flat and parallel to one another. (a: unknown, b: CM5408145, 89F)



SR-3: **Normal (a) and shingle (b)** It is not uncommon for an individual to both normal (a) and shingle ribs (b). (UT88-09, 82M)



SR-4: **Shingle**Both edges of the rib are extremely sharp (superior edge can be seen in the photograph). The distal third of the rib body is thin and flat (b), while the rest retains a more normal contour (a). (UT88-09, 82M)

Figure D1.13. R3-10 rib body thickness (shingle ribs) description and examples.

Clavicle medial epiphysis fusion [bilateral]

Location: edges and surface of the medial end of the clavicle

Scores: 0. not fused (NF) 1. partially fused (PF) 2. remnant line (RL) 3. fully fused (FF)

Not fused (NF): A ridged surface where there is no evidence of a fusing epiphysis.

Partially fused (PF): A low, smooth epiphysis covers part of the original ridged surface. It can be centrally or peripherally located.

Remnant line (RL): There is a narrow and shallow groove where the epiphysis joins the rest of the bone. The line is a shallow, groove at least 5 mm long with a fused bottom.

Fully fused (FF): There is no evidence of where the epiphysis fused to the bone.

Notes:

- When the medial clavicle exhibits a deep conical divot, which can be unilateral or bilateral, epiphyseal fusion is not scored.
- When a shallow or partial divot is present, the medial epiphysis can be scored, but a note should be made regarding the presence of the



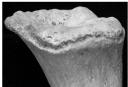
CEF-3: Partially fused (UP1058-2019, 20M)



CEF-4: Partially fused (CM110-57, 19F)



CEF-5: Partially fused (UP1369-5856, 27M)



CEF-6 Remnant line (NF-105)



CEF-7: Remnant line (UP1903-6390, 24F)



CEF-8: Remnant line (UP1965-6312, 27M)



CEF-1: Not fused (UT21-91, 25M)



CEF-2: Not fused (UT14-04, 19M)



r **a** ⁵-9: **Not scorable b** Interest (a) and oblique (b) views or a deep conical divot. The clavicle should not be scored when this feature is present. (UT23-94, 67M)

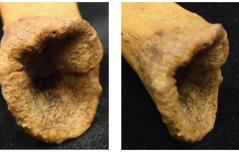


Figure D1.14. Clavicle medial epiphysis fusion description and examples.

Clavicle medial epiphysis gravel [bilateral]

Location: surface of the medical end of the clavicle

Scores: 0. absent 1. present

Present: A layer of irregular bone that is ≥ 5 mm in its longest dimension is present on top of the original clavicular surface. In many cases, the texture resembles a gravel road's surface.

Absent: The surface of the medial clavicle is smooth, with or without microporosity, or the irregular bone growth does not cover 5 mm or more of the surface. It the epiphysis is not present (not fused or partially fused), then the trait is scored as absent (e.g., CMEG-1, CMEG-2),

Notes:

- A shallow divot can change the texture of the medial epiphysis (see CMEG-3). Gravel typically exists on a roughened surface.
- The feature cannot be scored when a deep conical divot is present (e.g., CMEG-10).



CMEG-4:**Present** (UP1196-6411, 60M)



CMEG-5: **Present** (CM47-51, 81M)



CMEG-6: Present (CM22-51, 83F)



CMEG-1:**Absent** The epiphysis is unfused, so gravel is absent. (UT14-04, 19M)



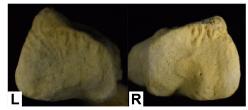
CMEG-2: **Absent** The epiphysis is only partially fused, so gravel is absent. (UP1058-2019, 20M)



CMEG-7:**Present** (UT41-05, 86M)



CMEG-8: **Present** (UP1629-6202, 70M)



CMEG-3:Absent
AbsentLeft and right clavicles with a common shape variant (shallow conical divot) that often results in a corrugated texture. This texture is regular and the bone is smooth. It should not be mistaken for irregular bone growth added to the surface. (CM122-50,36M)



CMEG-9: Present (UT15-09, 77F)



CMEG-10: **Not scorable** When a deep conical divot is present, the feature cannot be scored. (UT23-94, 67M)

Figure D1.15. Clavicle medial epiphysis gravel description and examples.

Clavicle medial macroporosity [bilateral]

Location: surface of the medical end of the clavicle

Scores: 0. absent 1. macroposity (≥3)

2. macroporosity (many)

Absent: The medial end is smooth, has only microporosity, or has fewer than three macropores.

It is also scored as absent when the epiphysis is not fused or partially fused (e.g., CMEM-1 & CMEM-2).

Macroporosity: The surface has 3 or more pores that are 1 mm or larger. The feature should be scored as many when more than 75% of the medial surface is macroporous.

Notes:

- Macropores can occur on the original surface
 of the bone or in (or around) areas of lumpy,
 irregular bone deposited on its surface. If
 present, the pores should be scored regardless
 of other textural changes.
- Developmental defects resembling macroporescan occur on the medial clavicular surface. These defects have rounded margins and smooth walls with fused bottoms that do notexpose the underlying trabecular bone (seeCMEM-3). These defects are similar to those sometimes found on other articulation surfaces, such as the iliac auricular surface.
- Macropores in the central part of a clavicle with a shallow or deep conical divot are not scored (e.g., CMEM-6).



CMEM-3: Absent Several surface defects are present in the central portion of a shallow conical divot. As opposed to true macropores, these defects occur on a dense surface and have rounded borders. (UT46-12, 38M)



CMEM-4: **Present** (CM31-52, 65F)



CMEM-5: **Present**Macroporosity is present
within irregular bone
growth. This is scored in the
same way as macropores
that occur directly on the
subcondral surface (e.g.,
CMEM-4).
(SA1629-6202, 70M)





CMEM-1:**Absent** The ephysis is unfused, so porosity is absent. (UT14-04, 19M)



CMEM-2: **Absent** The ephysis is only partially fused, so porosity is absent. (UP1058-2019, 20M)



CMEM-6: **Not scorable** When a deep conical divot is present, the feature typically cannot be scored. If the entire surface is visible, any macroporosity in the depths of the divot is not scored. (UT23-94, 67M)

Figure D1.16. Clavicle medial macroporosity description and examples.

Clavicle lateral macroporosity (acromial articulation) [bilateral]

Location: the lateral articulation surface where the clavicle meets the acromion process of the scapula

Scores: 0. absent 1. macroposity (≥3 holes)

Macroporosity: The articulation surface has 3 or more pores that are 1 mm or larger.

Absent: The articular surface is smooth, has only microporosity, or has fewer than three large pores.

Notes:

 Macropores must be on the normally smooth articulation surface to be scored. Macropores only on the edge of the bone should not be scored.

Examples:



CLM-3: **Present**Macropores are present within additional irregular bony growth.
(SA1629-6202, 70M)

Photo Not Yet Available

CLM-1: **Absent** Smooth.



CLM-4: **Present** (UT02-91, 81M)



CLM-2: **Absent**The surface has some microporosity, but no macropores are present. (CM122-50, 36M)



CLM-5: **Present** (UT13-04, 74M)

Figure D1.17. Clavicle lateral macroporosity description and examples.

Scapula glenoid fossa raised border (original definition) [bilateral]

Location: the margins of the articulation surface of the glenoid fossa

Scores: 0. absent 1. present (<1/2) 2. present (≥1/2)

Absent: The entire glenoid fossa has smooth rounded edges that flow without interruption from the articular surface to the external surface of the joint and beyond.

Present: The present category is coded by the amount of a raised rounded or sharp rim or lipping present. A rim is a distinctive raised and typically sharp to irregular narrow elevation at the margin of the glenoid cavity. Lipping refers to irregular bony growths much like what occur at the margins of other synovial joints. Lipping can either extend laterally (towards the humerus) or perpendicularly outwards from the joint.

Lipping is almost always easy to see. Bone should be examined carefully, however, for a simply elevated rim, which can be rounded and easily missed.

In the <1/2 category, some bony lipping is present but covers less than one-half of the circumference of glenoid fossa. In the ≥1/2 category, bony lipping is present around at least one-half of the circumference of the margin.



SGFS-1: **Absent**The borders of the entire glenoid fossa are rounded. No raised edges or lipping are present. (CM110-57, 19F)



SGFS-2: **Absent**The borders are more well-defined than (SGFS-1), but there is no raised edge or lipping present. (UT18-90, 27M)



SGFS-3: **Present (<1/2)** A portion of the margin is slightly raised (arrow) above the original surface of the articulation surface. (UT21-92, 25M)



SGFS-4: **Present (<1/2)** A small amound of lipping is present on the lower anterior margin (arrow). (UP1090-6401, 58M)



SGFS-5: **Present** (≥1/2) Minor lipping is present around the entire glenoid fossa. (UT78-08, 92F)



SGFS-6: **Present (≥1/2)**Lipping is present around almost all of the glenoid fossa. (UP1464-5665, 66M)



SGFS-7: **Present** (≥1/2) Lipping is present around almost all of the glenoid fossa. (CM53-50, 44M)



SGFS-8: **Present (≥1/2)**The entire glenoid fossa is eburnated and expanded with irregular lipping.
(CM17-51, 82M)

Figure D1.18. Scapula glenoid fossa raised border and lipping description and examples.

Humerus lesser tubercle bumps (anterior surface) [bilateral]

Location: anterior surface of the lesser tubercle of the humerus

Scores: 0. absent 1. present (≥1/3)

Present: Bumps range from low, rounded elevations just above the surface of the bone to large, irregular exostoses of several millimeters or more. Bumps must cover ≥1/3 of the anterior surface of the tubercle to be present.

Absent: The anterior surface of the lesser tubercle is smooth, or <1/3 of its surface is covered in bumps.

Notes:

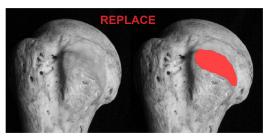
- Irregular pitting sometimes covers part or all ofthe anterior surface of the lesser tubercle.
 If theface is entirely covered in pitting, this featurecannot be scored. If some pitting is present, itshould be ignored, and the bone is scored basedon the criteria described above.
- This trait cannot be scored when eburnation and lipping of the humeral head has obliterated much or all of the lesser tubercle.



HLTS-1: **Absent**The anterior
surface of the
lesser tubercle is
completely smooth.
(UT18-90, 27M)



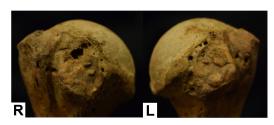
HLTS-2: **Absent**The anterior
surface of the
tubercle is slightly
textured, but no
irregular bone
growth (bumps) are
present that appear
as if they sit on
top of the surface.
(UT09-06, 35M)



HLTS Fig 1. Location of the anterior surface of the lesser tubercle of the humerus.



HLTS-3: **Present**Just over one-third
of the surface
is covered by
exostoses. The pits
are ignored.
Image reversed.
(UT05-92, 62F)



HLTS-4: **Present**Anterior views of the right and left less tubercles. The large pit on the superior portion of the right tubercle is ignored for scoring purposes. (UT55-04, 80F)



HLTS-5: **Present**Almost the entire surface of the tubercle is covered by rough, irregular, bony growth.
(UT37-09, 92F)



HLTS-6: **Present** (UT06-87, 69M)

Figure D1.19. Humerus lesser tubercle anterior surface bumps description and examples.

Humerus lesser tubercle margin shape [bilateral]

Location: lateral margin of the lesser tubercle of the humerus

Scores: 0. round/flat 1. raised 2. lipped

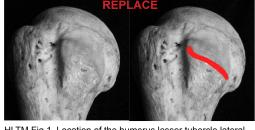
Round: The lateral margin and anterior surface of the tubercle are smooth and blend seamlessly together. The lateral margin of the lesser tubercle is rounded or flat, but not elevated above the tubercle's anterior surface.

Raised: At least half of the margin of the tubercle is raised above the anterior surface of the lesser tubercle, and this elevation can be seen as well as felt. The raised area can be either smooth or rough, but is clearly rises above the anterior surface of the tubercle.

Lipped: The margin of the tubercle is noticeably roughened with irregular bony growth extending above the rounded or shape tubercle margin. The margin is irregular and distinct from the tubercle's anterior face.



HLTM-1: **Round**The margin of the tubercle blends seamlessly into the anterior surface of the tubercle.
(UT09-06, 35M)



HLTM Fig 1. Location of the humerus lesser tubercle lateral margin.



HLTM-3: **Raised**The margin is raised, but no irregular bony growth is present.
Image reversed.
(CM 9-51, 73M)



HLTM-4: **Lipped**The margin is irregular and the anterior surface is smooth.
(CM 77-50, 62F)



HLTM-2: Round
The margin of the
tubercle is slightly
more defined than
in HMTM-1, but it
is not raised above
the anterior surface
of the tubercle.
(UT18-90, 27M)



HLTM-5: Lipped
The margin is covered in irregular bony growth, and the surface is covered in bumps. The margin and the surface are scored separately. (CM178-51, 87M)

Figure D1.20. Humerus lesser tubercle margin shape description and examples.

Humerus medial epicondyle [bilateral]

Location: anterior surface and posterior edge of the medial epicondyle of the humerus

Scores: 0. smooth

1. rough (≥1/2 exostoses or 5 mm rim)

Rough: Numerous small and flat exostoses produce a markedly rough surface that covers ≥1/2 of the epicondyle's anterior-medial surface or a rim at least 5 mm long is present on the posterior edge. Sometimes both exostoses and a rim are present, although the presence of one of the two features is sufficient for this feature to be considered rough.

Smooth: The surface and posterior border of this area is mostly smooth, without 5 mm of posterior rim or exostoses that cover ≥1/2 of the surface.

Notes:

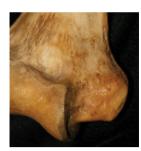
The rim has to be noticeably irregular lippingas if bone has been added to the edge ofthe epicondyle. A low, smooth raised bordercan occur, but it is classified as smooth.



HME-3 **Rough**Some exostoses, but
they do not cover onehalf of the surface. A
posterior rim more than
5 mm long is present.
(UT78-08, 92F)



HME-4: **Rough**Both a rim and
exostoses are present.
Image reversed.
(UT13-04, 74M)



HME-1: **Smooth**Anterior view of left
medial epicondyle with a
smooth anterior surface
and no posterior rim.
Image reversed.
(UT18-90, 27M)



HME-5: Rough Both a rim and exostsoses are present. Either feature is sufficient for the medial epicondyle to be scored as rough. Image reversed. (UT55-04, 80F)



HME-2: **Smooth**Anterior (a) and medial-oblique (b) views of the right medial epicondyle with a smooth anterior surface (a) and no posterior rim (b). In (b) a raised edge can be seen, but it is smooth, not irregular and rough. No lipping is present. (UP1348-5701, 76M)



HME-6: **Rough**Both a posterior rim and exostoses are present.
Image reversed. (UT05-92, 62F)

Figure D1.21. Humerus medial epicondyle description and examples.

Humerus lateral epicondyle [bilateral]

Location: anterior surface and posterior edge of the lateral epicondyle of the humerus

Scores: 0. smooth

1. rough (≥1/2 exostoses or 10 mm rim)

Rough: Numerous small and flat exostoses produce a markedly rough surface that covers ≥1/2 of the epicondyle's anterior-medial surface or a rim ≥10 mm long is present on the posterior edge. In some cases, both exostoses and a rim are present, although the presence of one of the two features is sufficient for this feature to be classified as rough.

Smooth: The surface and posterior border of this area is mostly smooth, without 5 mm of posterior rim or exostoses that cover ≥1/2 of the surface.

Notes:

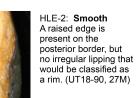
The exostoses and rim of the lateral epicondyle are often larger and more pronounced than those on the medial epicondyle.

Examples:

Photo Not Yet Available

Smooth surface with a rounded edge.

HLE-1: Smooth





HLE-3: Rough Surface exostoses with no rim.

Photo Not Yet

Available

Photo

Not Yet

Available

HLE-4: Rough Posterior rim with no exostoses.



HLE-5: Rough Both a posterior rim and exostoses are present. (UT05-92, 62F)



HLE-6: **Rough**Both a posterior rim and exostoses are present. (UT55-04, 80F)

Figure D1.22. Humerus lateral epicondyle description and examples.

Radial tuberosity medial crest [bilateral]

Location: medial aspect of the radial tuberosity where the tendon of biceps brachii inserts

Scores: 0. absent

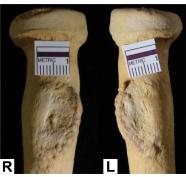
1. present (≥1 mm high and 10 mm long)

Present: A distinctive, narrow, and high crest of bone is present where the tendon inserts. This bony growth extends above the tuberosity surface by ≥1 mm for ≥10 mm in a proximal to distal direction. This ridge can have the appearance of a wave with its crest inclined over the normally smooth part of the tubercle where a bursa is located in life.

Absent: The medial edge of the tubercle is rounded, sharp, or only slightly lipped (<1 mm high) or the lipping extends for <10 mm in length.



RTMC-1: **Absent**L: The rounded and smooth edge of the tuberosity is not lipped. R: Slight lipping is present, but it does not extend above the tuberosity for ≥1 mm.
(UP1371-5221, 57M)



RTMC-2: **Absent**Both tuberosities are slighty raised and roughened, but there is insufficient lipping to be scored as present. (UP1448-5839, 86M)



RTMC-3: **Absent** Irregular lipping is present for ≥10 mm in length, but it does not extend off of the surface for ≥1 mm. (UP1034-5018, 59F)



RTMC-4: **Present**A borderline case, but just enough lipping is present
(≥1 mm high by 10 mm long. (UT21-06, 46M)



RTMC-5: **Present**Anterior (R1) and oblique (R2) views of the same right radius. The medial tuberosity has the distinctive wave-shaped crest. (UP1034-5018, 59M)



RTMC-6: Present Bilateral wave-shaped crests. (UP1464-5665, 60M)

Figure D1.23. Radial tuberosity medial crest description and examples.

Trapezium lipping [bilateral]

Location: the circumference of the saddle-shaped articulation surface for the first metacarpal

Scores: 0. absent 1. present (≥2 mm)

Present: At least 2 mm of lipping extends outward from at least two points along the margin of the saddle-shaped articulation surface. In advanced cases, the articulation surface becomes flat to irregular, lipping at the margin is extensive, and the bone no longer retains the overall shape of a trapezium (e.g., TL-#).

Absent: No lipping, or <2 mm of lipping, project outward from at least two points along the joint margin. The key feature is the absence of marked lipping. Sometimes bones marked as absent can display some eburnation, but there is little marginal lipping and the articulation surface retains its overall saddle-shaped appearance.

Notes:

 When the trapezium has fused to MC1 as aresult of arthritic bony lipping (not trauma), thefeature should be scored as present.

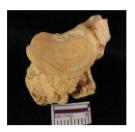
Examples:

Photo Not Yet Available

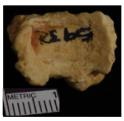
TL-1: **Absent**The margins are rounded. No lipping is present.



TL-2: **Absent**Only slight lipping is present around the left and right articular surfaces. (UP1166-5273, 66M)

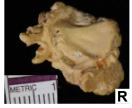


TL-3: **Absent**Approximately 1 mm of extra
bone growth is present around
the entire articular surface.
(UP1224-5587, 84M)



TL-4: **Present**More than 2 mm of bony
growth extends out from at
least two points around the
original articulation surface
margin. (UP1934-5982, 80M)





TL-5: **Present**More than 2 mm of lipping extends from a least two points on both the left and right bones. Eburnation is also present on the right side. (UP1889-6586, 83F)





TL-6: **Present**Bilateral lipped surfaces that also have extensive eburnation. (UP1186-6435, 85M)

Figure D1.24. Trapezium lipping description and examples.

Femur fovea margin lipping [bilateral]

Location: the margin of the femoral fovea where the ligamentum teres attaches (ligamentum capitis femoris)

Scores: 0. absent 1. lipped (≥10 mm)

Absent: The edge of the fovea is smooth or, more typically, it has a rounded or sharp margin.

Lipped: Irregular bony growths extend for ≥10 mm around the fovea margin. This irregular lipping can be as much as 2-4 mm wide, which looks like a flat plateau extending beyond the edge of the fovea.



FFML-3: **Lipped**The entire fovea is slightly irregular with some postmortem damage present on the inferior border.
(UP1334-5810, 70M)

Notes:

- In some cases, lipping around the fovea merges with an adjacent area of roughening on the femoral head (see Femoral Head Roughening, pg. 59).
- Rarely the fovea is small and appears filled in with bone with no margin clearly separable from the interior (e.g., NOT YET AVAILABLE). These atypical femora are not scored for this trait.



FFML-4: **Lipped**The entire fovea has a thick, irregular rim. (CM126-52, 58F)

Examples:



FFML-1 **Absent** (UP1152-5029,16F)



FFML-5: **Lipped** (UT100-06, 57F)



FFML-2 **Absent**The fovea has a raised and sharp, but regular, margin. For lipping to be present, the bone growth around the fovea must be irregular.



FFML-6: Unscorable The fovea cannot be scored because more than 10 mm of the fovea is damaged. It is not possible to tell how much lipping was originally present. (UP1492-3298, 31M)

Figure D1.25. Femur fovea margin lipping description and examples.

Femur head surface extra bone [bilateral]

Location: the femoral head, excluding the area in immediate proximity to the articulation margin with the frmoral neck (edge of the femoral head)

Scores: 0. absent 1. extra bone (5-9.9 mm) 2. extra bone (≥10 mm)

Extra bone: Extra bone measuring ≥5 mm in its longest dimension is present on the articulation surface.

Absent: The normally smooth articulation surface of the femoral head is either smooth, or has <5mm of a bone deposit. If pitting (removal of bone from the surface) is present, it should be ignored.

Notes:

- Sometimes lipping of the fovea margin and the rough femoral surface area join one another.
- Often the rough area lies immediately alongside the fovea, and it may join the lippingthat defines the fovea's edge. The bone deposits, however, can be islands of rough-looking bone separated by as much as a centimeter or more from the fovea. In the intervening area, the original smooth subchondral bone is visible.



FHEB-3: **Absent**The surface is slightly roughened but <5 mm of additional bone growth is present.
(UP1427-5319, 70F)



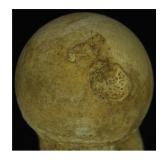
FHEB-4: Extra bone (≥ 5 mm)

The surface is slightly pitted, but more than 5mm of additional bony growth is present. (ADBOU - HOM1272 - X2573, A2541)





FHSR-1: **Absent** The entire femoral head is smooth. (UP1152-5029,16F)



FHEB-5: **Extra bone** (≥ **5 mm)** (UT03-96, 55F)



FHSR-2: **Absent** An area of pitting is present, but there is not more than 5 mm of extra bony growth. (UP1944-6265, 45M)



FHEB-6: **Extra bone** (≥ **10 mm)** (UP1121-5452, 84F)

Figure D1.26. Femur head surface extra bone growth description and examples.

Femur trochanteric fossa exostoses [bilateral]

Location: depths of the trochanteric fossa where the obturator externus inserts (FTFE Fig. 1)

Scores: 0. absent 1. present

Present: At least two exostoses are clearly visible in the depths of the fossa. Occasionally the exostoses will merge into a single large area with multiple peaks, and it is also scored as present.

Absent: The area is smooth with little or no evidence of generally sharp exostoses. A single exostosis is considered absent.

Examples:

Photo Not Yet Available

FTFE-1: **Absent** No exostoses are present.



FTFE-2: **Present** (UP1152-5029, 16F)



FTFE-3: **Present** Exostoses located slightly superior to the depths of the trochanteric fossa. Image reversed. (UT05-92, 62F)

Diagram Not Yet Available

FTFE Fig. 1: Trochanteric fossa - location of the obturator externus insertion





FTFE-4: **Present** L&R: Both the trochanteric fossa (circled) and medial trochanteric fossa exostsoses are present. R: The fossa exostoses have merged into a single large spike with multiple peaks. (UP1003-5073, 64F)







FTFE-6: **Present** Image reversed. (UP1121-5452, 84F)

Figure D1.27. Femur trochanteric fossa exostoses description and examples.

Femur trochanteric medial exostoses [bilateral]

Location: area immediately proximal to the trochanteric fossa on the medial side of the proximal end of the greater trochanter where obturator internus and gemelli insert

Scores: 0. absent 1. present

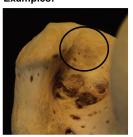
Present: A patch of ≥ 2 exostoses, similar to those seen in the trochanteric fossa, are present.

Absent: The area is smooth or has, at most, a single exostosis

Notes:

- The exostoses generally form a distinctly roughpatch of bone that is separate from the similarpatch of roughened bone deep within thetrochanteric fossa. The two areas areoccasionally confluent.
- Occasionally the exostoses take the form of a rough raised rim surrounding a small oval tendon insertion.
- In some individuals, this area has low raised, but regular, ridges that do not extend off the surface of the bone. In the absence of raised, distinct exostoses, the feature should be scored as

Examples:



FTFME-1: **Absent** (UP1152-5029, 16F)



FTFME-2: **Absent**Low rounded linear
elevations that are not
irregular additions of
bone. Compare these to
FTFME-6.
(UT25-06, 44F)

Diagram Not Yet Available

FTFE Fig. 1: Trochanteric fossa - location of the obturator externus insertion





FTFME-3: **Present** L&R: Both trochanteric fossa and medial trochanteric fossa (circled) exostsoses are present. R: The exostoses have merged into a single large spike with multiple peaks. R: Image reversed. (UP1003-5073, 64F)



FTFME-4: **Present** Image reversed. (UP1014-7049, 93M)



FTFME-5: **Present**Exostoses are almost merging with those in the trochanteric fossa. (UT13-04, 74M)

29

Figure D1.28. Femur medial trochanteric fossa exostoses description and examples (page 1 of 2)

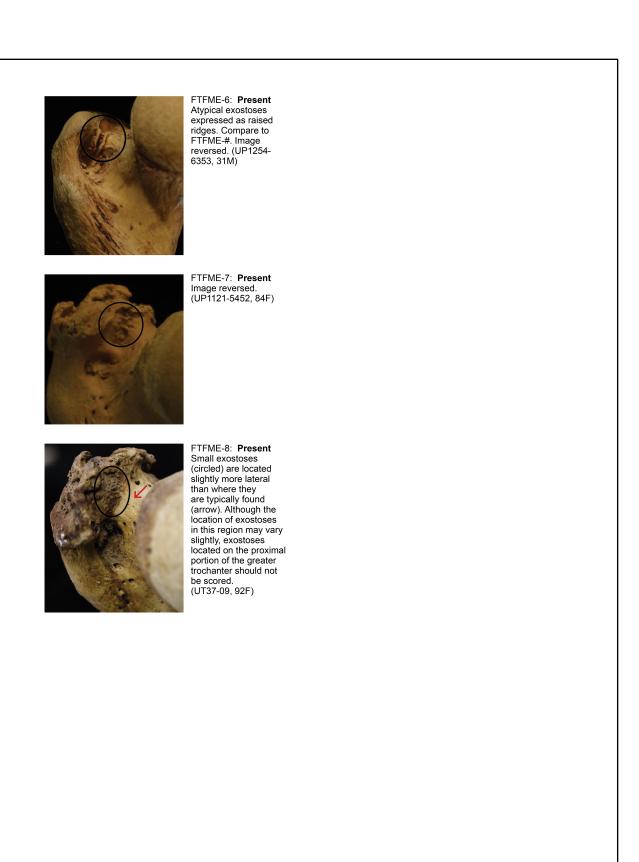


Figure D1.29. Femur medial trochanteric fossa exostoses description and examples (page 2 of 2).

Sacroiliac joint fusion [bilateral]

Location: the sacroiliac joint

Scores: 0. normal 1. fused 2. superior-anterior (SA)

Superior-anterior fusion (SA): The sacrum and ilium are fused via a bony connection in the region where the anterior sacroiliac ligament exists.

Fused: The sacrum is fused to the ilium through bony ossifications in any location. The location of the fusion should be noted.

Absent: If no bony union exists between the two elements, the trait is absent.

Notes:

If the sacrum, innominate, or both show damage in the region of the sacroiliac joint, they should be examined closely to determine if a bony connection between the two was broken postmortem (e.g., SAF-#).

Diagram Not Yet Available

SAF Fig. 1: Location of SA Fusion.



SAF-3: **Present**Bilateral superior-anterior fusion. (UT13-01, 86M)



SAF-1: **Present**Unilateral superior-anterior (SA) fusion with a small band of ossification (arrow). (UT02-87, 75M)



SAF-4: **Present** Bilateral superior-anterior fusion. (UT06-87, 69M)



SAF-2: **Present**Unilateral superior-anterior (SA) fusion with ossification spanning most of the superior portion of the joint.
(UP1160-5486, 76M)



SAF-5: **Present**Superior-anterior fusion that has been broken postmortem and rearticulated for the photo. Arrow indicates the location of the break. (UT32-09, 63M)

Figure D1.30. Superior-anterior sacroiliac joint fusion description and examples.

Ilium AllS exostoses (anterior inferior iliac spine exostoses) [bilateral]

Location: surface of the anterior inferior iliac spine

Scores: 0. absent (<75%) 1. exostoses (≥75%)

Exostoses: Numerous low rounded elevations of bone (exostoses) produce a rough surface that cover ≥75% of the anterior inferior iliac spine (AIIS).

Absent: The surface is smooth or, if exostoses are present, the roughened area covers <75% of the surface.

Notes:

 Small exostsoses can give the bone a slightly roughened appearance. If < 25% of the surface of the AIIS shows original, smooth, unmodified bone, the feature should be considered present (exostoses).

Examples:



AIIS-1: **Smooth** (UP1152-5029, 16F)

Photo Not Yet Available

AIIS-2: **Absent (<75%)** Some exostoses are present but they cover less than 75% of the area.

Diagram Not Yet Available

AIIS Fig. 1: Location of anterior inferior iliac spine (AIIS).



AIIS-3: **Present (≥75%)** (UP1121-5452, 84F)



AIIS-4: **Present (≥75%)** (UT07-08, 92F)

Figure D1.31. Ilium AIIS exostoses description and examples.

Acetabulum posterior margin lipping [bilateral]

Location: the 30 mm forming the posterior portion of the acetabulum's outer margin

Scores: 0. absent 1. present (≥10 mm)

Absent: The outer edge of the acetabulum is smooth and it can be either round or sharp. That is, in cross-section the rim is U- or V-shaped from the articulation (joint) surface to the outer surface of the bone.

Present: In the present category, noticeable and irregular bony lipping occurs along a continuous ≥10 mm of the acetabulum's margin.

Notes:

When lipping is present, traces of the original edge of the joint are often still visible. The lipping extends outward from the original edge. Frequently the easiest way to visualize the lipping is when viewing the bone from the acetabular surface.

Examples:



APML-1: **Absent** (UP1008-5885, 32F)

Photo Not Yet Available APML-2: **Absent**Posterior margin is sharp, but no lipping is present.

Diagram Not Yet Available

APML Fig. 1: Scoring location for posterior margin lipping.



APML-2: **Present (≥10 mm)** (UT05-92, 62F)



APML-3: **Present (≥10 mm)** Image reversed. (UP1384-5570, 83M)



APML-4: **Present (≥10 mm)** Image reversed. (CM119-49, 57F)

Figure D1.32. Acetabulum posterior margin lipping description and examples.

Acetabulum articular surface extra bone (posterior) [bilateral]

Location: the lateral posterior portion of the articular (joint) surface of the acetabulum, just inside the joint margin. The joint surface in the same region as the posterior acetabular margin described above

Scores: 0. absent 1. present (≥5 mm)

Present: An area of extra bone measuring ≥5 mm in its longest dimension is present on the articulation surface.

Absent: The normally smooth articulation surface of the acetabulum is either smooth, or has <5 mm of additional bone growth. If pitting (removal of bone from the surface) is present, it should be ignored.

Notes:

 The irregular bony growth is typically separated from the joint margin, and any lipping that might occur there, by a narrow strip of the original joint surface. However, there are occasions when the margin lipping and joint surface roughening are immediately adjacent to one other.

Examples:



APMR-1: **Absent** (UP1008-5885, 32F)



APMR-2: **Absent**Additional bone is present on the articular surface, but it is smooth and regular. This is not the irregular bony growth characteristic of this trait.
Compare with APMR-#.
(UP1380-5316, 41F)

Diagram Not Yet Available

APMR Fig. 1: Scoring location for posterior margin lipping.

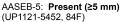


AASEB-4: Present (≥5 mm)

Image reversed. (CM140-57, 89M)

AASEB-3: **Present (≥5 mm)** Image reversed. (CM13-50, 72M)







AASEB-6: **Present (≥5 mm)** UP1629-6202, 70M)

Figure D1.33. Acetabulum articular surface extra bone growth description and examples.

Ischial tuberosity bumps [bilateral]

Location: upper portion of the ischial tuberosity; this area is superior to a low ridge that separates the attachments for the semitendinosus and semimembranosus superiorly, and the long-head of the biceps femoris inferiorly

Scores: 0. absent 1. present (<50%) 2. present (≥50%)

Present: Bumpy, irregular bony growths disfigure the shape of the originally smooth ischial tuberosity. In the present <50% category, at least one bump is present. Present ≥50% means that most of the surface is covered by low bony growths.

Absent: This portion of the ischial tuberosity is smooth.

Examples:



ITB-1: **Absent** No irregular ossifications are present. (UT21-92, 25M)



ITB-2: **Smooth**The surface is slightly more coarse than ITB-#, but no irregular ossifications are present. (UP1008-5885, 32F)



ITB-3: **Present (<50%)** (UT02-07, 81M)



ITB-4: **Present (<50%)** A single ossification is present. (CM160-52, 46F)

Diagram Not Yet Available

ITB Fig. 1: Scoring location for ischial tuberosity bumps.



ITB-5: Present (≥50%)
Borderline case, but just over 50% of the surface area is covered with additional bone growth. Irregular bone added to the surface ranges from small bumps with the texture of coarse sand to lumps that are significantly raised off of the originally articulation surface. (UP1121-5452, 84F)



ITB-6: **Present** (≥50%) More than one-half of the scoring area is covered in dense, irregular ossifications. (UT55-04, 80F)



ITB-7: **Present (≥50%)**The entire surface of the ischial tuberosity is covered in irregular ossifications. (UT41-05, 86F)

Figure D1.34. Ischial tuberosity bumps description and examples.

Humerus weight

Location: the entire humerus

Scores: 0. normal 1. light

Normal: The bone weight falls within what one might expect for an individual of that size. Bones that are borderline, or only slightly lighter than normal, should be scored as normal. This observation is subjective, so err on the side of classifying the bone as normal.

Light: The bone must be noticeably lighter than expected, but otherwise normal in appearance. A light humerus typically feels like an extremely light, hollow, cortical bone shell.

Notes:

- This trait should be recorded or, at least, mentally noted when the bones are first picked up, usually as the box is being unpacked. Handling bones for some time before scoring can result in skewed perceptions of relative bone weight.
- In rare cases bone weight is asymmetric. When this occurs, record both sides, and make a note on the data form.

Tibia weight

Location: the entire tibia

Scores: 0. normal 1. light

Normal: The bone weight falls within what one might expect for an individual of that size. Bones that are borderline, or only slightly lighter than normal, should be scored as normal. This observation is subjective, so err on the side of classifying the bone as normal.

Light: The bone must be noticeably lighter than expected, but otherwise normal in appearance. A light tibia is often little more than a cortical bone shell.

Notes:

- This trait should be recorded or, at least, mentally noted, when the bones are first picked up, usually when unpacking the skeleton's box. Handling bones for some time before scoring can result in skewed perceptions of relative bone weight.
- In only rare cases will bone weight be asymmetric. When that occurs, record both sides and make a note on the data collection

Calcaneus weight

Location: the entire calcaneus

Scores: 0. normal 1. light

Normal: The bone weight falls within what one might expect for an individual of that size. Bones that are borderline, or only slightly lighter than normal, should be scored as normal. This observation is subjective, so err on the side of classifying the bone as normal.

Light: The bone must be noticeably lighter than expected. The cortical bone is often extremely fragile and paper-thin.

Notes:

- This trait should be recorded or, at least, mentally noted, when the bones are first picked up, usually as the skeleton's box is unpacked. Handling bones for some time before scoring can result in skewed perceptions of relative bone weight.
- In only rare cases will bone weight be asymmetric. When that occurs, record both sides and make a note on the data form.

Innominate weight

Location: the entire innominate

Scores: 0. normal 1. light

Normal: The bone weight falls within what one might expect for an individual of that size. Bones that are borderline, or only slightly lighter than normal, should be scored as normal. This observation is subjective, so err on the side of classifying the bone as normal.

Light: The bone must be noticeably lighter than expected, but otherwise normal in appearance. In light innominates, the cortical bone is often extremely fragile and paper-thin.

Notes:

- This trait should be recorded or, at least, mentally noted, when the bones are first picked up, usually as the skeleton's box is unpacked. Handling bones for some time before scoring can result in skewed perceptions of relative bone weight.
- In only rare cases will bone weight be asymmetric. When that occurs, record both sides and make a note on the data form.

Figure D1.35. Subjective bone weight descriptions.

APPENDIX E: CORRELATION MATRICES FOR PHASE 3 TRAITS

This appendix contains correlation matrices for all pairs of the 40 features used in Phase 3. The correlations are based on the Phase 1 reference data set (N=1010). A coefficient of one means perfect agreement, while zero indicates no relationship. A positive coefficient indicates that when one trait is present the other is also likely to occur, with a higher number indicating a greater probability of the traits both appearing in the same individual. A negative coefficient indicates the opposite relationship, where if one trait is present, the other is less likely to occur. Coefficients are colored by the strength of the relationship (yellow: 0.3–0.499, orange: 0.5–0.599, red: 0.6–0.99) to aid in assessing patterns although the extent to which the magnitude of the correlation will impact age estimates is unknown.

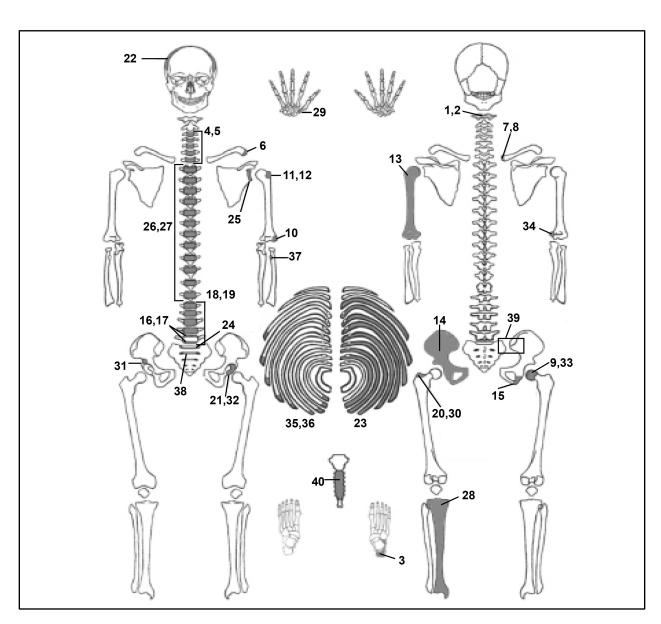


Figure E1.0 Locations of the features in Phase 3 trait correlation matrices (Tables E1.0-E1.7).

Table E1.0. Correlation matrix for traits 1-6

- 1001	E I.U. Correlation matrix to	1	2	3	4	5	6
	Trait	C1E	C1Lip	CalWT	Cwax	Clip	ClavLatMacro
1	C1E		0.357	0.180	0.060	0.335	0.240
2	C1Lip	0.357		0.142	0.076	0.308	0.275
3	CalWT	0.180	0.142		0.009	0.130	0.128
4	CWax	0.060	0.076	0.009		0.133	0.065
5	CLip	0.335	0.308	0.130	0.133		0.327
6	ClavLatMacro	0.240	0.275	0.128	0.065	0.327	
7	ClavFus	0.080	0.166	0.041	0.016	0.133	0.125
8	ClavMedGrav	0.251	0.259	0.171	0.141	0.372	0.328
9	ForMar	0.253	0.288	0.172	0.045	0.337	0.276
10	HumLatEpi	0.262	0.365	0.144	0.100	0.406	0.363
11	HumLesTubBumps	0.323	0.220	0.142	0.093	0.325	0.259
12	HumLesTubLip	0.410	0.360	0.192	0.129	0.466	0.398
13	HumWT	0.167	0.102	0.482	0.024	0.128	0.126
14	OSWT	0.240	0.129	0.541	-0.018	0.103	0.132
15	IschTubBumps	0.177	0.141	0.216	0.107	0.191	0.193
16	L5InfMarLip	0.265	0.375	0.169	0.076	0.434	0.378
17	L5SupMarLip	0.187	0.358	0.138	0.063	0.393	0.343
18	LWax	0.100	0.065	0.011	0.218	0.103	0.045
19	Llip	0.280	0.289	0.112	0.132	0.480	0.341
20	FemMedTroch	0.292	0.259	0.101	0.063	0.361	0.269
21	AcePostMar	0.275	0.303	0.205	0.078	0.356	0.268
22	ParDep	0.114	0.043	0.089	-0.021	0.042	0.025
23	R310Shingle	0.130	0.041	0.201	-0.026	0.082	0.008
24	S1SupRnd	0.168	0.268	0.119	0.051	0.280	0.277
25	ScapFossaRnd	0.167	0.369	0.107	0.049	0.313	0.286
26	Twax	0.217	0.154	0.017	0.188	0.268	0.199
27	TLip	0.305	0.301	0.121	0.091	0.393	0.298
28	TibWT	0.156	0.101	0.627	0.021	0.109	0.107
29	TrapLip	0.244	0.242	0.177	0.055	0.263	0.174
30	TrochFos	0.268	0.272	0.121	0.067	0.342	0.296
31	AIIS	0.265	0.199	0.225	0.109	0.305	0.202
32	AcePostSur	0.160	0.143	0.136	0.056	0.158	0.096
33	FemHead	0.121	0.136	0.097	0.050	0.198	0.166
34	HumMedEpi	0.316	0.284	0.118	0.130	0.385	0.292
35	R2Edge	0.257	0.273	0.107	0.084	0.313	0.274
36	R310Edge	0.237	0.356	0.141	0.082	0.428	0.312
37	RadCrest	0.253	0.187	0.033	0.115	0.252	0.203
38	S12 Fusion	0.088	0.212	0.052	0.023	0.177	0.175
39	SIFusion	0.106	0.068	-0.010	0.130	0.153	0.139
40	SternRidge	0.223	0.198	0.306	0.115	0.210	0.159

Table E1.1. Correlation matrix for traits 7-11

	e ET.T. Correlation matrix	7	8	9	10	11
	Trait	ClavFus	ClavMedGrav	ForMar	HumLatEpi	HumLesTubBumps
1	C1E	0.080	0.251	0.253	0.262	0.323
2	C1Lip	0.166	0.259	0.288	0.365	0.220
3	CalWT	0.041	0.171	0.172	0.144	0.142
4	CWax	0.016	0.141	0.045	0.100	0.093
5	CLip	0.133	0.372	0.337	0.406	0.325
6	ClavLatMacro	0.125	0.328	0.276	0.363	0.259
7	ClavFus		NA	0.168	0.152	0.068
8	ClavMedGrav	NA		0.376	0.353	0.454
9	ForMar	0.168	0.376		0.417	0.296
10	HumLatEpi	0.152	0.353	0.417		0.349
11	HumLesTubBumps	0.068	0.454	0.296	0.349	
12	HumLesTubLip	0.123	0.489	0.408	0.520	0.541
13	HumWT	0.027	0.116	0.112	0.169	0.171
14	OSWT	0.031	0.158	0.147	0.142	0.158
15	IschTubBumps	0.050	0.250	0.212	0.245	0.321
16	L5InfMarLip	0.228	0.371	0.444	0.498	0.312
17	L5SupMarLip	0.284	0.288	0.392	0.451	0.243
18	LWax	0.017	0.156	0.073	0.092	0.183
19	Llip	0.126	0.349	0.292	0.412	0.369
20	FemMedTroch	0.113	0.311	0.380	0.332	0.354
21	AcePostMar	0.089	0.361	0.376	0.349	0.324
22	ParDep	0.012	-0.006	0.055	0.062	0.081
23	R310Shingle	0.018	0.089	0.099	0.081	0.044
24	S1SupRnd	0.264	0.244	0.401	0.366	0.213
25	ScapFossaRnd	0.355	0.239	0.423	0.349	0.178
26	Twax	0.045	0.358	0.209	0.261	0.320
27	TLip	0.115	0.429	0.333	0.400	0.347
28	TibWT	0.027	0.147	0.143	0.154	0.154
29	TrapLip	0.049	0.188	0.136	0.217	0.223
30	TrochFos	0.112	0.360	0.452	0.405	0.331
31	AIIS	0.070	0.323	0.284	0.272	0.358
32	AcePostSur	0.027	0.180	0.154	0.134	0.120
33	FemHead	0.043	0.178	0.248	0.137	0.203
34	HumMedEpi	0.092	0.398	0.301	0.519	0.388
35	R2Edge	0.222	0.267	0.355	0.285	0.278
36	R310Edge	0.261	0.297	0.447	0.419	0.309
37	RadCrest	0.064	0.273	0.192	0.301	0.346
38	S12 FusionOpen	0.561	0.095	0.192	0.193	0.094
39	SIFusion	0.031	0.200	0.146	0.115	0.190
40	SternRidge	0.069	0.252	0.301	0.266	0.274

Table E1.2. Correlation matrix for traits 12-16

Table	E 1.2. Correlation matrix i	12	13	14	15	16
	Trait	HumLesTubLip	HumWT	OSWT	IschTubBumps	L5InfMarLip
1	C1E	0.410	0.167	0.240	0.177	0.265
2	C1Lip	0.360	0.102	0.129	0.141	0.375
3	CalWT	0.192	0.482	0.541	0.216	0.169
4	CWax	0.129	0.024	-0.018	0.107	0.076
5	CLip	0.466	0.128	0.103	0.191	0.434
6	ClavLatMacro	0.398	0.126	0.132	0.193	0.378
7	ClavEpiphFus	0.123	0.027	0.031	0.050	0.228
8	ClavMedGravel	0.489	0.116	0.158	0.250	0.371
9	ForMar	0.408	0.112	0.147	0.212	0.444
10	HumLatEpi	0.520	0.169	0.142	0.245	0.498
11	HumLesTubBumps	0.541	0.171	0.158	0.321	0.312
12	HumLesTubLip		0.155	0.207	0.354	0.510
13	HumWT	0.155		0.625	0.254	0.112
14	OSWT	0.207	0.625		0.263	0.157
15	IschTubBumps	0.354	0.254	0.263		0.218
16	L5InfMarLip	0.510	0.112	0.157	0.218	
17	L5SupMarLip	0.403	0.107	0.117	0.176	0.625
18	LWax	0.112	0.024	0.044	0.082	0.054
19	Llip	0.458	0.103	0.110	0.307	0.487
20	FemMedTroch	0.426	0.123	0.104	0.258	0.341
21	AcePostMar	0.445	0.178	0.192	0.246	0.390
22	ParDep	0.086	0.120	0.126	0.001	0.062
23	R310Shingle	0.094	0.193	0.212	0.141	0.062
24	S1SupRound	0.317	0.086	0.091	0.171	0.593
25	ScapFossaRound	0.329	0.084	0.097	0.151	0.502
26	Twax	0.307	0.089	0.096	0.193	0.193
27	TLip	0.442	0.122	0.153	0.281	0.369
28	TibWT	0.163	0.709	0.608	0.248	0.099
29	TrapLip	0.300	0.205	0.176	0.195	0.195
30	TrochFos	0.480	0.120	0.138	0.218	0.381
31	AIIS	0.420	0.144	0.147	0.326	0.267
32	AcePostSur	0.136	0.176	0.156	0.111	0.137
33	FemHead	0.190	0.071	0.015	0.139	0.151
34	HumMedEpi	0.505	0.117	0.139	0.241	0.398
35	R2Edge	0.396	0.138	0.130	0.186	0.296
36	R310Edge	0.412	0.119	0.151	0.204	0.427
37	RadCrest	0.315	0.087	0.048	0.189	0.237
38	S12 Fusion	0.149	0.039	0.047	0.076	0.237
39	SIFusion	0.189	0.006	-0.010	0.017	0.136
40	SternRidge	0.324	0.335	0.306	0.388	0.286

Table E1.3. Correlation matrix for traits 17-22

	le E1.3. Correlation matri	17	18	19	20	21	22
	Trait	L5SupMarLip	LWax	LLip	FemMedTroch	AcePostMar	ParDep
1	C1E	0.187	0.100	0.280	0.292	0.275	0.114
2	C1Lip	0.358	0.065	0.289	0.259	0.303	0.043
3	CalWT	0.138	0.011	0.112	0.101	0.205	0.089
4	CWax	0.063	0.218	0.132	0.063	0.078	-0.021
5	CLip	0.393	0.103	0.480	0.361	0.356	0.042
6	ClavLatMacro	0.343	0.045	0.341	0.269	0.268	0.025
7	ClavFus	0.284	0.017	0.126	0.113	0.089	0.012
8	ClavMedGravel	0.288	0.156	0.349	0.311	0.361	-0.006
9	ForMar	0.392	0.073	0.292	0.380	0.376	0.055
10	HumLatEpi	0.451	0.092	0.412	0.332	0.349	0.062
11	HumLesTubBumps	0.243	0.183	0.369	0.354	0.324	0.081
12	HumLesTubLip	0.403	0.112	0.458	0.426	0.445	0.086
13	HumWT	0.107	0.024	0.103	0.123	0.178	0.120
14	OSWT	0.117	0.044	0.110	0.104	0.192	0.126
15	IschTubBumps	0.176	0.082	0.307	0.258	0.246	0.001
16	L5InfMarLip	0.625	0.054	0.487	0.341	0.390	0.062
17	L5SupMarLip		0.064	0.435	0.325	0.342	0.052
18	LWax	0.064		0.139	0.056	0.060	0.058
19	Llip	0.435	0.139		0.300	0.378	0.046
20	FemMedTroch	0.325	0.056	0.300		0.342	-0.008
21	AcePostMar	0.342	0.060	0.378	0.342		0.069
22	ParDep	0.052	0.058	0.046	-0.008	0.069	
23	R310Shingle	0.046	0.078	0.005	0.123	0.116	0.143
24	S1SupRnd	0.523	0.055	0.366	0.220	0.281	0.043
25	ScapFossaRnd	0.475	0.049	0.295	0.301	0.263	0.046
26	Twax	0.172	0.302	0.320	0.195	0.233	0.039
27	TLip	0.361	0.132	0.503	0.277	0.361	0.052
28	TibWT	0.100	0.021	0.063	0.124	0.184	0.125
29	TrapLip	0.181	0.059	0.245	0.170	0.225	0.066
30	TrochFos	0.334	0.080	0.359	0.537	0.336	-0.018
31	AIIS	0.219	0.071	0.319	0.342	0.354	0.060
32	AcePostSur	0.097	0.024	0.156	0.084	0.209	0.058
33	FemHead	0.122	0.022	0.166	0.204	0.226	0.007
34	HumMedEpi	0.316	0.127	0.382	0.220	0.277	0.000
35	R2Edge	0.335	0.065	0.270	0.281	0.339	-0.006
36	R310Edge	0.489	0.042	0.334	0.300	0.377	0.063
37	RadCrest	0.219	0.116	0.307	0.226	0.209	0.025
38	S12 Fusion	0.235	0.023	0.154	0.145	0.130	0.018
39	SIFusion	0.103	0.223	0.162	0.161	0.119	-0.041
40	SternRidge	0.245	0.051	0.227	0.276	0.297	0.018

Table E1.4. Correlation matrix for traits 23-27

iable	E1.4. Correlation matrix to	23	24	25	26	27
	Trait	R310Single	S1SupRnd	ScapFossaRnd	Twax	TLip
1	C1E	0.130	0.168	0.167	0.217	0.305
2	C1Lip	0.041	0.268	0.369	0.154	0.301
3	CalWT	0.201	0.119	0.107	0.017	0.121
4	CWax	-0.026	0.051	0.049	0.188	0.091
5	CLip	0.082	0.280	0.313	0.268	0.393
6	ClavLatMacro	0.008	0.277	0.286	0.199	0.298
7	ClavFus	0.018	0.264	0.355	0.045	0.115
8	ClavMedGrav	0.089	0.244	0.239	0.358	0.429
9	ForMar	0.099	0.401	0.423	0.209	0.333
10	HumLatEpi	0.081	0.366	0.349	0.261	0.400
11	HumLesTubBumps	0.044	0.213	0.178	0.320	0.347
12	HumLesTubLip	0.094	0.317	0.329	0.307	0.442
13	HumWT	0.193	0.086	0.084	0.089	0.122
14	OSWT	0.212	0.091	0.097	0.096	0.153
15	IschTubBumps	0.141	0.171	0.151	0.193	0.281
16	L5InfMarLip	0.062	0.593	0.502	0.193	0.369
17	L5SupMarLip	0.046	0.523	0.475	0.172	0.361
18	LWax	0.078	0.055	0.049	0.302	0.132
19	Llip	0.005	0.366	0.295	0.320	0.503
20	FemMedTroch	0.123	0.220	0.301	0.195	0.277
21	AcePostMar	0.116	0.281	0.263	0.233	0.361
22	ParDep	0.143	0.043	0.046	0.039	0.052
23	R310Shingle		0.024	0.057	-0.008	0.042
24	S1SupRnd	0.024		0.432	0.132	0.293
25	ScapFossaRnd	0.057	0.432		0.134	0.292
26	Twax	-0.008	0.132	0.134		0.402
27	TLip	0.042	0.293	0.292	0.402	
28	TibWT	0.207	0.094	0.084	0.038	0.088
29	TrapLip	0.102	0.139	0.121	0.179	0.224
30	TrochFos	0.085	0.317	0.297	0.234	0.321
31	AIIS	0.133	0.211	0.209	0.269	0.294
32	AcePostSur	0.054	0.083	0.084	0.141	0.121
33	FemHead	0.133	0.080	0.127	0.090	0.127
34	HumMedEpi	0.033	0.284	0.242	0.311	0.392
35	R2Edge	0.104	0.250	0.304	0.156	0.320
36	R310Edge	0.093	0.348	0.378	0.168	0.317
37	RadCrest	0.028	0.196	0.175	0.253	0.268
38	S12 Fusion	0.028	0.250	0.390	0.062	0.155
39	SIFusion	-0.047	0.101	0.089	0.293	0.190
40	SternRidge	0.126	0.225	0.196	0.176	0.276

Table E1.5. Correlation matrix for traits 28-32

		28	29	30	31	32
	Trait	TibWT	TrapLip	TrochFos	AIIS	AcePostSur
1	C1E	0.156	0.244	0.268	0.265	0.160
2	C1Lip	0.101	0.242	0.272	0.199	0.143
3	CalWT	0.627	0.177	0.121	0.225	0.136
4	CWax	0.021	0.055	0.067	0.109	0.056
5	CLip	0.109	0.263	0.342	0.305	0.158
6	ClavLatMacro	0.107	0.174	0.296	0.202	0.096
7	ClavFus	0.027	0.049	0.112	0.070	0.027
8	ClavMedGrav	0.147	0.188	0.360	0.323	0.180
9	ForMar	0.143	0.136	0.452	0.284	0.154
10	HumLatEpi	0.154	0.217	0.405	0.272	0.134
11	HumLesTubBumps	0.154	0.223	0.331	0.358	0.120
12	HumLesTubLip	0.163	0.300	0.480	0.420	0.136
13	HumWT	0.709	0.205	0.120	0.144	0.176
14	OSWT	0.608	0.176	0.138	0.147	0.156
15	IschTubBumps	0.248	0.195	0.218	0.326	0.111
16	L5InfMarLip	0.099	0.195	0.381	0.267	0.137
17	L5SupMarLip	0.100	0.181	0.334	0.219	0.097
18	LWax	0.021	0.059	0.080	0.071	0.024
19	Llip	0.063	0.245	0.359	0.319	0.156
20	FemMedTroch	0.124	0.170	0.537	0.342	0.084
21	AcePostMar	0.184	0.225	0.336	0.354	0.209
22	ParDep	0.125	0.066	-0.018	0.060	0.058
23	R310Shingle	0.207	0.102	0.085	0.133	0.054
24	S1SupRnd	0.094	0.139	0.317	0.211	0.083
25	ScapFossaRnd	0.084	0.121	0.297	0.209	0.084
26	Twax	0.038	0.179	0.234	0.269	0.141
27	TLip	0.088	0.224	0.321	0.294	0.121
28	TibWT		0.174	0.128	0.177	0.220
29	TrapLip	0.174		0.146	0.098	0.125
30	TrochFos	0.128	0.146		0.284	0.127
31	AIIS	0.177	0.098	0.284		0.137
32	AcePostSur	0.220	0.125	0.127	0.137	
33	FemHead	0.069	0.131	0.196	0.154	0.100
34	HumMedEpi	0.127	0.228	0.300	0.281	0.116
35	R2Edge	0.111	0.125	0.273	0.200	0.100
36	R310Edge	0.132	0.194	0.324	0.250	0.113
37	RadCrest	0.095	0.108	0.215	0.226	0.128
38	S12 Fusion	0.040	0.068	0.152	0.101	0.041
39	SIFusion	-0.015	-0.010	0.138	0.151	0.043
40	SternRidge	0.369	0.257	0.227	0.224	0.161

Table E1.6. Correlation matrix for traits 33-36

		33	34	35	36
	Trait	FemHead	HumMedEpi	R2Edge	R310Edge
1	C1E	0.121	0.316	0.257	0.237
2	C1Lip	0.136	0.284	0.273	0.356
3	CalWT	0.097	0.118	0.107	0.141
4	CWax	0.050	0.130	0.084	0.082
5	CLip	0.198	0.385	0.313	0.428
6	ClavLatMacro	0.166	0.292	0.274	0.312
7	ClavFus	0.043	0.092	0.222	0.261
8	ClavMedGrav	0.178	0.398	0.267	0.297
9	ForMar	0.248	0.301	0.355	0.447
10	HumLatEpi	0.137	0.519	0.285	0.419
11	HumLesTubBumps	0.203	0.388	0.277	0.309
12	HumLesTubLip	0.190	0.505	0.396	0.412
13	HumWT	0.071	0.117	0.138	0.119
14	OSWT	0.015	0.139	0.130	0.151
15	IschTubBumps	0.139	0.241	0.186	0.204
16	L5InfMarLip	0.151	0.398	0.296	0.427
17	L5SupMarLip	0.122	0.316	0.335	0.489
18	LWax	0.022	0.127	0.065	0.042
19	Llip	0.166	0.382	0.270	0.334
20	FemMedTroch	0.204	0.220	0.281	0.300
21	AcePostMar	0.226	0.277	0.339	0.377
22	ParDep	0.007	0.000	-0.006	0.063
23	R310Shingle	0.133	0.033	0.104	0.093
24	S1SupRnd	0.080	0.284	0.250	0.348
25	ScapFossaRnd	0.127	0.242	0.304	0.378
26	Twax	0.090	0.311	0.156	0.168
27	TLip	0.127	0.392	0.320	0.317
28	TibWT	0.069	0.127	0.111	0.132
29	TrapLip	0.131	0.228	0.125	0.194
30	TrochFos	0.196	0.300	0.273	0.324
31	AIIS	0.154	0.281	0.200	0.250
32	AcePostSur	0.100	0.116	0.100	0.113
33	FemHead		0.123	0.190	0.183
34	HumMedEpi	0.123		0.264	0.320
35	R2Edge	0.190	0.264		0.743
36	R310Edge	0.183	0.320	0.743	
37	RadCrest	0.064	0.359	0.272	0.226
38	S12 Fusion	0.065	0.125	0.253	0.291
39	SIFusion	0.042	0.158	0.123	0.077
	SternRidge	0.160	0.330	0.239	0.205

Table E1.7. Correlation matrix for traits 37-40

		37	38	39	40
	Trait	RadCrest	S12Fusion	SIFusion	SternRidge
1	C1E	0.253	0.088	0.106	0.223
2	C1Lip	0.187	0.212	0.068	0.198
3	CalWT	0.033	0.052	-0.010	0.306
4	CWax	0.115	0.023	0.130	0.115
5	CLip	0.252	0.177	0.153	0.210
6	ClavLatMacro	0.203	0.175	0.139	0.159
7	ClavFus	0.064	0.561	0.031	0.069
8	ClavMedGrav	0.273	0.095	0.200	0.252
9	ForMar	0.192	0.192	0.146	0.301
10	HumLatEpi	0.301	0.193	0.115	0.266
11	HumLesTubBumps	0.346	0.094	0.190	0.274
12	HumLesTubLip	0.315	0.149	0.189	0.324
13	HumWT	0.087	0.039	0.006	0.335
14	OSWT	0.048	0.047	-0.010	0.306
15	IschTubBumps	0.189	0.076	0.017	0.388
16	L5InfMarLip	0.237	0.237	0.136	0.286
17	L5SupMarLip	0.219	0.235	0.103	0.245
18	LWax	0.116	0.023	0.223	0.051
19	Llip	0.307	0.154	0.162	0.227
20	FemMedTroch	0.226	0.145	0.161	0.276
21	AcePostMar	0.209	0.130	0.119	0.297
22	ParDep	0.025	0.018	-0.041	0.018
23	R310Shingle	0.028	0.028	-0.047	0.126
24	S1SupRnd	0.196	0.250	0.101	0.225
25	ScapFossaRnd	0.175	0.390	0.089	0.196
26	Twax	0.253	0.062	0.293	0.176
27	TLip	0.268	0.155	0.190	0.276
28	TibWT	0.095	0.040	-0.015	0.369
29	TrapLip	0.108	0.068	-0.010	0.257
30	TrochFos	0.215	0.152	0.138	0.227
31	AIIS	0.226	0.101	0.151	0.224
32	AcePostSur	0.128	0.041	0.043	0.161
33	FemHead	0.064	0.065	0.042	0.160
34	HumMedEpi	0.359	0.125	0.158	0.330
35	R2Edge	0.272	0.253	0.123	0.239
36	R310Edge	0.226	0.291	0.077	0.205
37	RadCrest		0.088	0.146	0.207
38	S12 Fusion	0.088		0.044	0.084
39	SIFusion	0.146	0.044		-0.032
	SternRidge	0.207	0.084	-0.032	

REFERENCES

- Abbie, A. (1950). Closure of cranial articulations in the skull of the Australian aborigine. *Journal of anatomy, 84*(Pt 1), 1.
- Acsadi, J., & Nemeskeri, G. (1970). History of human life span and mortality. *Académiai Kiadó, Budapest*.
- Agresti, A. (2007a). Generalized Linear Models *An Introduction to Categorical Data Analysis* (2nd ed., pp. 65-98). Hoboken, NJ: John Wiley and Sons.
- Agresti, A. (2007b). Logistic Regression *An Introduction to Categorical Data Analysis* (2md ed., pp. 99-136). Hoboken, NJ: John Wiley and Sons.
- Aiello, L. C., & Molleson, T. (1993). Are microscopic ageing techniques more accurate than macroscopic ageing techniques? *Journal of archaeological science*, *20*(6), 689-704.
- Aktas, E. Ö., Koçak, A., Aktas, S., & Yemisçigil, A. (2004). Intercostal Variation for Age Estimation—Are the Standards for the Right 4th Rib Applicable for Other Ribs? *Collegium Antropologicum*, 28(2), 267-272.
- Albert, A. M., & Maples, W. R. (1995). Stages of epiphyseal union for thoracic and lumbar vertebral centra as a method of age determination for teenage and young adult skeletons. *Journal of Forensic Science*, 40(4), 623-633.
- Alesbury, H. S., Ubelaker, D. H., & Bernstein, R. (2013). Utility of the frontonasal suture for estimating age at death in human skeletal remains. *Journal of Forensic Sciences*, *58*(1), 104-108.
- Algee-Hewitt, B. F., Tersigni-Tarrant, M., & Shirley, N. (2013). Age estimation in modern forensic anthropology. *Forensic Anthropology: An Introduction*, 181-219.
- Anderson, H. (2007). Towns. In M. Carver & J. Klapste (Eds.), *The Archaeology of Medieval Europe, Vol. 2: Twelfth to Sixteenth Centuries* (Vol. 2, pp. 370-278). Aarhus, Denmark: Aarhus University Press.
- Anderson, M. F., Anderson, D. T., & Wescott, D. J. (2010). Estimation of adult skeletal age-at-death using the Sugeno fuzzy integral. *American Journal of Physical Anthropology*, 142(1), 30-41.
- Angel, J. L. (1971). *The People of Lerna: Analysis of a Prehistoric Aegean Population* (Vol. II). Washington, DC: Smithsonia Institution Press.
- Angel, J. L., Suchey, J. M., Iscan, M. Y., & Zimmerman, M. R. (1986). Age at death estimated from the skeleton and viscera. *Dating and Age Determination of Biological Materials. Croom Helm, London*, 179-220.
- Apostolidou, C., Eleminiadis, I., Koletsa, T., Natsis, K., Dalampiras, S., Psaroulis, D., . . . Njau, S. (2011). Application of the maxillary suture obliteration method for estimating age at death in Greek population. *Open Forensic Science Journal*, *4*, 15-19.

- Archives of Ontario. (2012a). Living Longer, Living Healthier: Health Education in the Curative Age, 1921-1947. Retrieved from http://www.archives.gov.on.ca/en/explore/online/health_promotion/living_longer.aspx
- Archives of Ontario. (2012b). Sanitation, public hygiene, and the fight against disease: 1882 to 1921. Retrieved from http://www.archives.gov.on.ca/en/explore/online/health_promotion/sanitation_hygiene_disease.aspx
- Arking, R. (2006a). Measuring Age-related Changes in Populations *The Biology of Aging: Observation and Principles* (pp. 26-53). New York, NY: Oxford University Press.
- Arking, R. (2006b). Perspectives on Aging *The Biology of Aging: Observations and Principles* (pp. 3-25). New York, NY: Oxford University Press.
- Armstrong, G. L., Conn, L. A., & Pinner, R. W. (1999). Trends in infectious disease mortality in the united states during the 20th century. *JAMA*, 281(1), 61-66. doi:10.1001/jama.281.1.61
- Austad, S. N. (1994). Menopause: An evolutionary perspective. *Experimental Gerontology*, 29(3-4), 255-263.
- Baccino, E., Ubelaker, D. H., Hayek, L.-A. C., & Zerilli, A. (1999). Evaluation of seven methods of estimating age at death from mature human skeletal remains. *Journal of Forensic Science*, *44*(5), 931-936.
- Baker, R. K. (1984). The relationship of cranial suture closure and age analyzed in a modern multi-racial sample of males and females. (MA Thesis Microfilm), California State University, Fullerton, CA.
- Barchilon, V., Hershkovitz, I., Rothschild, B., Wish-Baratz, S., Latimer, B., Jellema, L., . . . Arensburg, B. (1996). Factors affecting the rate and pattern of the first costal cartilage ossification. *The American journal of forensic medicine and pathology, 17*(3), 239-247.
- Barrier, P., Dedouit, F., Braga, J., Joffre, F., Rougé, D., Rousseau, H., & Telmon, N. (2009). Age at death estimation using multislice computed tomography reconstructions of the posterior pelvis. *Journal of Forensic Sciences*, *54*(4), 773-778.
- Beauthier, J. P., Lefevre, P., Meunier, M., Orban, R., Polet, C., Werquin, J. P., & Quatrehomme, G. (2010). Palatine sutures as age indicator: a controlled study in the elderly. *Journal of Forensic Sciences*, *55*(1), 153-158.
- Bedford, M., Russell, K., & Lovejoy, C. (1989). The auricular surface aging technique. 16 color photographs with descriptions. *Kent, Ohio: Kent State University*.
- Bedford, M., Russell, K. F., Lovejoy, C. O., Meindl, R. S., Simpson, S. W., & Stuart-Macadam, P. L. (1993). Test of the multifactorial aging method using skeletons with known ages-at-death from the Grant collection. *American Journal of Physical Anthropology*, *91*(3), 287-297.

- Bednarek, J., Bloch-Bogusławska, E., & Sliwka, K. (2001). Use of morphologic changes in the pubic symphysis for age determination in the Polish male population. *Archiwum medycyny sadowej i kryminologii*, *52*(4), 295-304.
- Beise, J. (2005). The Helping and the Helpful Grandmother: The Role of Maternal and Paternal Grandmothers in Child Mortality in the Seventeenth- and Eighteenth-Century Population of French Settlers in Quebec, Canada. In E. Voland, A. Chasiotis, & W. Schiefenhovel (Eds.), *Grandmotherhood* (pp. 215-238). New Brunswick, NJ: Rutgers University Press.
- Belkin, V., Livshits, G., Otremski, I., & Kobyliansky, E. (1998). Aging bone score and climatic factors. *American Journal of Physical Anthropology*, *106*(3), 349-359.
- Benedictow, O. J. (2008). Demographic conditions *Cambridge History of Scandinavia* (Vol. 1, pp. 235-249). Online: Cambridge University Press.
- Berg, G. E. (2008). Pubic bone age estimation in adult women. *Journal of Forensic Sciences*, 53(3), 569-577.
- Binstock, R. H. (1991). From the Great Society to the Aging Society--25 Years of the Older Americans Act. *Generations*, *15*(3), 11-18.
- Black, S., & Scheuer, L. (1996). Age changes in the clavicle: from the early neonatal period to skeletal maturity. *International Journal of Osteoarchaeology*, *6*(5), 425-434.
- Blanc, A., Rougé, D., Joffre, F., Telmon, N., Chemla, P., & Gaston, A. (2005). Application of the Suchey-Brooks method to three-dimensional imaging of the pubic symphysis. *Journal of Forensic Science*, *50*(3), 1-6.
- Bocquet-Appel, J.-P., & Masset, C. (1982). Farewell to paleodemography. *Journal of Human Evolution*, 11(4), 321-333.
- Bocquet-Appel, J.-P., & Masset, C. (1985). Paleodemography: resurrection or ghost? *Journal of Human Evolution*, *14*(2), 107-111.
- Boldsen, J. L. (2009). The Demography of Disease in Antiquity Disentangling the Effect of Latency Period and the Disease Related Risk of Dying. Paper presented at the Baltic Bioarchaeology Meeting, Vilnius, Lithuania.
- Boldsen, J. L., Milner, G. R., Konigsberg, L. W., & Wood, J. W. (2002). Transition analysis: a new method for estimating age from skeletons. In R. D. Hoppa & J. W. Vaupel (Eds.), *Paleodemography: age distributions from skeletal samples* (pp. 73-106). Cambridge, UK: Cambridge University Press.
- Bongiovanni, R. (2016). Effects of Parturition on Pelvic Age Indicators. *Journal of Forensic Sciences*.
- Boyd, K. L., Villa, C., & Lynnerup, N. (2015). The use of CT scans in estimating age at death by examining the extent of ectocranial suture closure. *Journal of Forensic Sciences*, *60*(2), 363-369.

- Broca, P. (1875). *Instructions craniologiques et craniométriques* (Vol. 2). Paris: Librairie Georges Masson Place De L'Ecole-De-Médecine.
- Brooks, S. T. (1955). Skeletal age at death: the reliability of cranial and pubic age indicators. *American Journal of Physical Anthropology*, *13*(4), 567-597.
- Brooks, S. T., & Suchey, J. M. (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, J. L. (2015). The (mis) use of adult age estimates in osteology. *Annals of human biology*, *42*(4), 323-331.
- Buckberry, J. L., & Chamberlain, A. T. (2002). Age estimation from the auricular surface of the ilium: a revised method. *American Journal of Physical Anthropology*, 119(3), 231-239.
- Buikstra, J. E., & Konigsberg, L. W. (1985). Paleodemography: critiques and controversies. *American Anthropologist*, *87*(2), 316-333.
- Buk, Z., Kordik, P., Bruzek, J., Schmitt, A., & 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), 294. e291-294. e299.
- Bullock, M., Márquez, L., Hernández, P., & Ruíz, F. (2013). Paleodemographic age-at-death distributions of two Mexican skeletal collections: A comparison of transition analysis and traditional aging methods. *American Journal of Physical Anthropology*, *152*(1), 67-78.
- Calce, S. E. (2012). A new method to estimate adult age-at-death using the acetabulum. American Journal of Physical Anthropology, 148(1), 11-23.
- Campanacho, V., Santos, A. L., & Cardoso, H. F. (2012). 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. *American Journal of Physical Anthropology*, *148*(3), 371-378.
- Cannon, A. (1995). Material Culture and Burial Representativeness. In S. R. Saunders & A. Herring (Eds.), *Grave Reflections: Portraying the Past Through Cemetery Studies*. Toronto, CA: Canadian Scholars' Press.
- Cant, M. A., & Johnstone, R. A. (2008). Reproductive conflict and the separation of reproductive generations in humans. *Proceedings of the National Academy of Sciences, 105*(14), 5332-5336.
- Cant, M. A., Johnstone, R. A., & Russel, A. F. (2009). Reproductive conflict and the evolution of menopause. In R. Hager & C. B. Jones (Eds.), *Reproductive skew in vertebrates:* proximate and ultimate causes (pp. 24–50): Cambridge University Press.
- Cardoso, A. F., & Henderson, C. Y. (2010). Enthesopathy Formation in the Humerus: Data from Known Age-at-Death and Known Occupation Skeletal Collections. *American Journal of Physical Anthropology*, *141*, 550-560.

- Cardoso, H. F. (2006). Brief communication: the collection of identified human skeletons housed at the Bocage Museum (National Museum of Natural History), Lisbon, Portugal. *American Journal of Physical Anthropology, 129*(2), 173-176.
- Centers for Disease Control and Prevention. (1999). Impact of vaccines universally recommended for children--United States, 1990-1998. MMWR. Morbidity and mortality weekly report, 48(12), 243.
- Cerezo-Román, J. I., & Espinoza, P. O. H. (2014). Estimating age at death using the sternal end of the fourth ribs from Mexican males. *Forensic science international*, 236, 196. e191-196. e196.
- Chamberlain, A. T. (2006). *Demography in archaeology*: Cambridge University Press.
- Chen, X., Zhang, Z., & Tao, L. (2008). Determination of male age at death in Chinese Han population: Using quantitative variables statistical analysis from pubic bones. *Forensic science international*, 175(1), 36-43.
- Chen, X., Zhang, Z., Zhu, G., & Tao, L. (2011). Determining the age at death of females in the Chinese Han population: Using quantitative variables and statistical analysis from pubic bones. *Forensic science international*, *210*(1), 278. e271-278. e278.
- Chiba, F., Makino, Y., Motomura, A., Inokuchi, G., Torimitsu, S., Ishii, N., . . . Yajima, D. (2013). Age estimation by multidetector CT images of the sagittal suture. *International journal of legal medicine*, *127*(5), 1005-1011.
- Coale, A., & Demeny, P. (1983). Regional model life tables and stable populations. New York, NY: Academic Press.
- Colarusso, T. (2016). A Test of the Passalacqua Age at Death Estimation Method Using the Sacrum. *Journal of Forensic Sciences*, *61*(S1).
- Colgrove, J. (2006). State of Immunity: The Politics of Vaccination in Twentieth-Century America. Berkeley: University of California Press.
- Corsini, M.-M., Schmitt, A., & 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), 163-167.
- Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*: CRC Press.
- Cox, M. (2000). Ageing adults from the skeleton. *Human osteology in archaeology and forensic science*, 61-82.
- Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The determinants of mortality. *The Journal of Economic Perspectives*, 20(3), 97-120.
- Cutler, D., & Miller, G. (2005). The role of public health improvements in health advances: the twentieth-century United States. *Demography*, *42*(1), 1-22.

- Dahlbäck, G. (2008). The towns. In K. Helle (Ed.), *The Cambridge History of Scandinavia:* Volume 1: Prehistory to 1520 (Vol. 1, pp. 611-634). Online: Combridge University Press.
- Davies, H. M. (1913). A consideration of the influence of the first costal cartilage on apical tuberculosis, based on a study of 402 specimens. *British Journal of Surgery*, 1(1), 55-69.
- Debono, L., Mafart, B., Guipert, G., & Jeusel, É. (2004). Application pratique de la méthode d'estimation de l'âge au décès de Schmitt et Broqua (2000). *Bulletins et Mémoires de la Société d'Anthropologie de Paris*(16 (1-2)), 115-120.
- Dedouit, F., Bindel, S., Gainza, D., Blanc, A., Joffre, F., Rougé, D., & Telmon, N. (2008). Application of the Iscan Method to Two-and Three-Dimensional Imaging of the Sternal End of the Right Fourth Rib. *Journal of Forensic Sciences*, *53*(2), 288-295.
- DeWitte, S. N. (2010). Age patterns of mortality during the Black Death in London, AD 1349–1350. *Journal of archaeological science*, *37*(12), 3394-3400.
- DiGangi, E. A., Bethard, J. D., Kimmerle, E. H., & Konigsberg, L. W. (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., & Marinković, J. (2007). Evaluation of the Suchey–Brooks method for aging skeletons in the Balkans. *Journal of Forensic Sciences*, *52*(1), 21-23.
- Dokladal, M. (1975). Cranial sutures obliteration and teeth attrition. Scripta Medica, 48, 225-240.
- Dorandeu, A., Coulibaly, B., Piercecchi-Marti, M.-D., Bartoli, C., Gaudart, J., Baccino, E., & Leonetti, G. (2008). Age-at-death estimation based on the study of frontosphenoidal sutures. *Forensic science international*. *177*(1), 47-51.
- Dorandeu, A., de la Grandmaison, G. L., Coulibaly, B., Durigon, M., Piercecchi-Marti, M.-D., Baccino, E., & Leonetti, G. (2009). Value of histological study in the fronto-sphenoidal suture for the age estimation at the time of death. *Forensic science international, 191*(1), 64-69.
- Dudar, J. C. (1993). Identification of rib number and assessment of intercostal variation at the sternal rib end. *Journal of Forensic Science*, *38*(4), 788-797.
- Dudar, J. C., Pfeiffer, S., & Saunders, S. R. (1993). Evaluation of morphological and histological adult skeletal age-at-death estimation techniques using ribs. *Journal of Forensic Science*, *38*(3), 677-685.
- Dudzik, B., & Langley, N. R. (2015). Estimating age from the pubic symphysis: A new component-based system. *Forensic science international*, *257*, 98-105.
- Dwight, T. (1890a). The Closure of the Cranial Sutures as a Sign of Age. *Boston Medical and Surgical Journal*, *122*(17), 389-392.
- Dwight, T. (1890b). Sternum as an index of sex, height, and age. *Journal of anatomy and physiology*, 24(4), 527-535.

- Eliopoulos, C., Lagia, A., & Manolis, S. (2007). A modern, documented human skeletal collection from Greece. *HOMO-Journal of Comparative Human Biology*, *58*(3), 221-228.
- Elkeles, A. (1966). Sex differences in the calcification of the costal cartilages. *Journal of the American Geriatrics Society*, *14*(5), 456-462.
- Fairgrieve, S. I., & Oost, T. S. (1995). On a test of the multifactorial aging method by Bedford et al.(1993). *American Journal of Physical Anthropology*, *97*(1), 83-85.
- Falys, C. G., & Lewis, M. E. (2011). Proposing a way forward: a review of standardisation in the use of age categories and ageing techniques in osteological analysis (2004–2009). *International Journal of Osteoarchaeology*, *21*(6), 704-716.
- Falys, C. G., & Prangle, D. (2015). Estimating age of mature adults from the degeneration of the sternal end of the clavicle. *American Journal of Physical Anthropology*, *156*(2), 203-214.
- Falys, C. G., Schutkowski, H., & Weston, D. A. (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.
- Fanton, L., Gustin, M. P., Paultre, U., Schrag, B., & Malicier, D. (2010). Critical study of observation of the sternal end of the right 4th rib. *Journal of Forensic Sciences*, *55*(2), 467-472.
- Faraway, J. J. (2006a). Likelihood Theory Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. New York, NY: Chapman & Hall/CRC.
- Faraway, J. J. (2006b). Nonparametric Regression *Extending the Linear Model with R:*Generalized Linear, Mixed Effects and Nonparametric Regression Models. New York, NY: Chapman & Hall/CRC.
- Felson, D. T., Lawrence, R. C., Dieppe, P. A., Hirsch, R., Helmick, C. G., Jordan, J. M., . . . Fries, J. F. (2000). Osteoarthritis: New Insights. Part 1: The Disease and Its Risk Factors. *Ann Intern Med, 133*(8), 635-646.
- Ferrant, O., Rougé-Maillart, C., Guittet, L., Papin, F., Clin, B., Fau, G., & Telmon, N. (2009). Age at death estimation of adult males using coxal bone and CT scan: a preliminary study. *Forensic science international, 186*(1), 14-21.
- Fleischman, J. M. (2013). A Comparative Assessment of the Chen et al. and Suchey-Brooks Pubic Aging Methods on a North American Sample. *Journal of Forensic Sciences*, *58*(2), 311-323.
- Forensic Anthropology Center. (2012). WM Bass Donated Skeletal Collection. Retrieved from https://fac.utk.edu/wm-bass-donated-skeletal-collection/
- Fou, A. A. (1997). John H. Gibbon. The first 20 years of the heart-lung machine. *Texas Heart Institute Journal*, *24*(1), 1.

- Franklin, D. (2010). Forensic age estimation in human skeletal remains: current concepts and future directions. *Legal Medicine*, 12(1), 1-7.
- Galera, V., Ubelaker, D. H., & Hayek, L.-A. C. (1995). Interobserver error in macroscopic methods of estimating age at death from the human skeleton. *International Journal of Anthropology*, *10*(4), 229-239.
- Galera, V., Ubelaker, D. H., & Hayek, L.-A. C. (1998). Comparison of macroscopic cranial methods of age estimation applied to skeletons from the Terry Collection. *Journal of Forensic Science*, *43*(5), 933-939.
- Garamendi, P. M., Landa, M. I., Botella, M. C., & Alemán, I. (2011). Forensic Age Estimation on Digital X-ray Images: Medial Epiphyses of the Clavicle and First Rib Ossification in Relation to Chronological Age. *Journal of Forensic Sciences*, *56*(s1).
- Garvin, H. M., & Passalacqua, N. V. (2012). Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences*, *57*(2), 427-433.
- Gauthier, J., & Schutkowski, H. (2013). Assessing the application of tooth cementum annulation relative to macroscopic aging techniques in an archeological sample. *HOMO-Journal of Comparative Human Biology*, 64(1), 42-57.
- Gavrilova, N. S., & Gavrilov, L. A. (2005). Human Longevity and Reproduction: An Evolutionary Perspective. In E. Voland, A. Chasiotis, & W. Schiefenhovel (Eds.), *Grandmotherhood* (pp. 59-80). New Brunswick, NJ: Rutgers University Press.
- Gilbert, B. M. (1973). Misapplication to females of the standard for aging the male os pubis. *American Journal of Physical Anthropology*, *38*(1), 39-40.
- Gilbert, B. M., & McKern, T. W. (1973). A method for aging the female os pubis. *American Journal of Physical Anthropology*, *38*(1), 31-38.
- Ginter, J. K. (2005). A test of the effectiveness of the revised maxillary suture obliteration method in estimating adult age at death. *Journal of Forensic Science*, *50*(6), JFS2004520-2004527.
- Gissel, S. (1981). The late medieval agrarian crisis in Denmark. In N. Skyum-Nielsen & N. Lund (Eds.), *Danish Medieval History, New Currents* (pp. 238-250.). Copenhagen, SK: Museum Tusculanum Press.
- Gocha, T. P., Ingvoldstad, M. E., Kolatorowicz, A., Cosgriff-Hernandez, M.-T. J., & Sciulli, P. W. (2015). Testing the applicability of six macroscopic skeletal aging techniques on a modern Southeast Asian sample. *Forensic science international, 249*, 318. e311-318. e317.
- Godde, K., & Hens, S. M. (2012). Age-at-death estimation in an Italian historical sample: A test of the Suchey-Brooks and transition analysis methods. *American Journal of Physical Anthropology*, *149*(2), 259-265.

- Godde, K., & Hens, S. M. (2015). Modeling senescence changes of the pubic symphysis in historic Italian populations: a comparison of the Rostock and forensic approaches to aging using transition analysis. *American Journal of Physical Anthropology, 156*(3), 466-473.
- Google (Cartographer). (2016). [Map of Horsens, Denmark at four different scales]. Retrieved from https://goo.gl/maps/B9e63gnuYzC2
- Graves, W. W. (1922). Observations on age changes in the scapula. A preliminary note. *American Journal of Physical Anthropology, 5*(1), 21-33.
- Gruspier, K. L., & Mullen, G. J. (1991). Maxillary suture obliteration: a test of the Mann method. *Journal of Forensic Science*, *36*(2), 512-519.
- Gupta, P., Rai, H., Kalsey, G., & Gargi, J. (2007). Age determination from sternal ends of the Ribs-an autopsy study. *Journal of Indian Academy of Forensic Medicine*, *29*(4), 94-97.
- Guyer, B., Freedman, M. A., Strobino, D. M., & Sondik, E. J. (2000). Annual summary of vital statistics: trends in the health of Americans during the 20th century. *Pediatrics*, *106*(6), 1307-1317.
- Hanihara, K., & Suzuki, T. (1978). Estimation of age from the pubic symphysis by means of multiple regression analysis. *American Journal of Physical Anthropology, 48*(2), 233-239.
- Hannallah, D., White, A. P., Goldberg, G., & Albert, T. J. (2007). Diffuse idiopathic skeletal hyperostosis. *Operative Techniques in Orthopaedics*, *17*(3), 174-177.
- Harth, S., Obert, M., Ramsthaler, F., Reuß, C., Traupe, H., & Verhoff, M. A. (2009). Estimating age by assessing the ossification degree of cranial sutures with the aid of Flat-Panel-CT. *Legal Medicine*, *11*, S186-S189.
- Harth, S., Obert, M., Ramsthaler, F., Reuß, C., Traupe, H., & Verhoff, M. A. (2010). Ossification Degrees of Cranial Sutures Determined with Flat-Panel Computed Tomography:
 Narrowing the Age Estimate with Extrema. *Journal of Forensic Sciences*, *55*(3), 690-694.
- Hartnett, K. M. (2010a). Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample—Part I: Pubic Bone. *Journal of Forensic Sciences*, *55*(5), 1145-1151.
- Hartnett, K. M. (2010b). 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.
- Harvey, W. (1968). Some Dental and Social Conditions of 1696–1852 Connected with St. Bride's Church, Fleet Stree, London. *Medical history*, *12*(01), 62-75.
- Hawkes, K., & Blurton Jones, N. (2005). Human Age Structures, Pleodemography, and the Grandmother Hypothesis. In E. Voland, A. Chasiotis, & W. Schiefenhovel (Eds.), *Grandmotherhood: the evolutionary significance of the second half of female life* (pp. 118-140). New Brunswick, NJ: Rutgers University Press.

- Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (1997). Hadza women's time allocation, offspring provisioning, and the evolution of long postmenopausal life spans. *Current Anthropology*, *38*(4), 551-577.
- Hawkes, K., O'Connell, J. F., Jones, N. B., Alvarez, H., & Charnov, E. L. (1998).

 Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences*, *95*(3), 1336-1339.
- Hens, S. M., & Belcastro, M. G. (2012). Auricular surface aging: a blind test of the revised method on historic Italians from Sardinia. *Forensic science international, 214*(1), 209. e201-209. e205.
- Hens, S. M., & Godde, K. (2016). Auricular Surface Aging: Comparing Two Methods that Assess Morphological Change in the Ilium with Bayesian Analyses. *Journal of Forensic Sciences*, *61*(S1).
- Hens, S. M., Rastelli, E., & 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.
- Hershkovitz, I., Latimer, B., Dutour, O., Jellema, L. M., Wish-Baratz, S., Rothschild, C., & Rothschild, B. M. (1997). Why do we fail in aging the skull from the sagittal suture? *American Journal of Physical Anthropology*, *103*(3), 393-399.
- Hill, K. R., & Hurtado, A. M. (1996). *Ache life history: The ecology and demography of a foraging people*: Transaction Publishers.
- Holman, D. J., O'Connor, K. A., & Wood, J. W. (2006). Age and Female Reproductive Function: Identifying the Most Important Biological Determinants. *Human Clocks: The Bio-cultural Meanings of Age, 5*, 171.
- Hoppa, R. D. (2000). Population variation in osteological aging criteria: an example from the pubic symphysis. *American Journal of Physical Anthropology, 111*(2), 185-191.
- Hoppa, R. D., & Saunders, S. (1998). The MAD legacy: How meaningful is mean age-at-death in skeletal samples. *Human evolution*, *13*(1), 1-14.
- Hoppa, R. D., & Vaupel, J. (2002). The Rostock Manifesto for paleodemography: the way from stage to age. *CAMBRIDGE STUDIES IN BIOLOGICAL AND EVOLUTIONARY ANTHROPOLOGY*, 1-8.
- Horiuchi, S., & Wilmoth, J. R. (1998). Deceleration in the Age Pattern of Mortality at Older Ages. *Demography, 35*(4), 391-412.
- Houck, M. M., Ubelaker, D. H., Owsley, D., Craig, E., Grant, W., Fram, R., . . . Sandness, K. (1996). The role of forensic anthropology in the recovery and analysis of Branch Davidian compound victims: assessing the accuracy of age estimations. *Journal of Forensic Science*, *41*(5), 796-801.
- Howden, L. M., & Meyer, J. A. (2011). Age and sex composition: 2010. 2010 Census Briefs, US Department of Commerce, Economics and Statistics Administration. US CENSUS BUREAU.

- Howell, N. (1976). Toward a uniformitarian theory of human paleodemography. *Journal of Human Evolution*, *5*(1), 25-40.
- Howell, N. (1979). Demography of the Dobe! kung.
- Howell, N. (1982). Village composition implied by a paleodemographic life table: the Libben site. *American Journal of Physical Anthropology*, *59*(3), 263-269.
- Hunt, D., & Albanese, J. (2005). History and demographic composition of the Robert J. Terry anatomical collection. *American Journal of Physical Anthropology*, *127*, 406-417.
- Hurst, C. V. (2010). A Test of the Forensic Application of Transition Analysis With the Pubic Symphysis. In K. E. Latham & J. M. Finnegan (Eds.), *Age estimation of the human skeleton* (pp. 262-272). Springfield, IL: Charles C Thomas.
- Igarashi, Y., Uesu, K., Wakebe, T., & Kanazawa, E. (2005). New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology*, 128(2), 324-339.
- Işcan, M. Y. (1989). Assessment of Age at Death in the Human Skeleton. In M. Y. Işcan (Ed.), *Age markers in the human skeleton* (pp. 5-18). Speingfield, IL: Charles C. Thomas.
- Işcan, M. Y. (1991). The aging process in the rib: An analysis of sex-and race-related morphological variation. *American Journal of Human Biology, 3*(6), 617-623.
- İşcan, M. Y., & Loth, S. R. (1986a). Determination of age from the sternal rib in white males: a test of the phase method. *Journal of Forensic Science*, *31*(1), 122-132.
- İşcan, M. Y., & Loth, S. R. (1986b). Estimation of age and determination of sex from the sternal rib *Forensic Osteology: Advances in the identification of human remains* (pp. 68-89). Springfield, IL: Charles C Thomas.
- İşcan, M. Y., & Loth, S. R. (1989). Osteological manifestations of age in the adult. Reconstruction of Life from the Skeleton, 23-40.
- İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984a). Age estimation from the rib by phase analysis: white males. *Journal of Forensic Science*, 29(4), 1094-1104.
- İşcan, M. Y., Loth, S. R., & Wright, R. K. (1984b). Metamorphosis at the sternal rib end: a new method to estimate age at death in white males. *American Journal of Physical Anthropology*, *65*(2), 147-156.
- İşcan, M. Y., Loth, S. R., & Wright, R. K. (1985). Age estimation from the rib by phase analysis: white females. *Journal of Forensic Science*, *30*(3), 853-863.
- İşcan, M. Y., Wright, R. K., & Loth, S. R. (1987). Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Science*, 32(2), 452-466.
- Jackes, M. (1985). Pubic symphysis age distributions. *American Journal of Physical Anthropology*, 68(2), 281-299.

- Jackes, M. (2000). Building the bases for paleodemographic analysis: adult age determination. In M. A. Katzenberg & S. Saunders (Eds.), *Biological Anthropology of the Human Skeleton* (pp. 417-466). New York, NY: Wiley-Liss.
- Jantz, R. L., & Ousley, S. D. (2013). Introduction to Fordisc 3. In M. A. Tersigni-Tarrant & N. R. Shirley (Eds.), *Forensic Anthropology: An Introduction* (pp. 253-270). Boca Raton, FL: CRC Press.
- Johansen, H. C. (2002). *Danish Population History 1600 1939*. Odense, DK: Hans Christian Andersen Center, University of Southern Denmark, University Press of Southern Denmark.
- Johnson, J. (1976). A comparison of age estimation using discriminant function analysis with some other age estimations of unknown skulls. *Journal of anatomy*, *121*(Pt 3), 475.
- Johnston, F. E., & Snow, C. E. (1961). The reassessment of the age and sex of the Indian Knoll skeletal population: demographic and methodological aspects. *American Journal of Physical Anthropology*, *19*(3), 237-244.
- Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology Issues News and Reviews*, *9*(4), 156-185.
- Katz, D., & Suchey, J. M. (1986). Age determination of the male os pubis. *American Journal of Physical Anthropology*, 69(4), 427-435.
- Katz, D., & Suchey, J. M. (1989). Race differences in pubic symphyseal aging patterns in the male. *American Journal of Physical Anthropology*, *80*(2), 167-172.
- Key, C. A., Aiello, L. C., & Molleson, T. (1994). Cranial suture closure and its implications for age estimation. *International Journal of Osteoarchaeology*, *4*(3), 193-207.
- Khandare, S., Bhise, S., & Shinde, A. (2015). Age estimation from cranial sutures—a Postmortem study. *International J. of Healthcare and Biomedical Research, 3*(03), 192-202.
- Khazaeli, A. A., Xiu, L., & Curtsinger, J. W. (1995). Stress Experiments as a Means of Investigating Age-Specific Mortality in Drosophila melongaster. *Experimental Gerontology*, 30, 177-184.
- Kimmerle, E. H., Konigsberg, L. W., Jantz, R. L., & Baraybar, J. P. (2008). Analysis of Age-at-Death Estimation Through the Use of Pubic Symphyseal Data. *Journal of Forensic Sciences*, *53*(3), 558-568.
- Kimmerle, E. H., Prince, D. A., & Berg, G. E. (2008). Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth. *Journal of Forensic Sciences*, 53(3), 594-600.
- King, J. B. (1939). Calcification of the costal cartilages. *The British Journal of Radiology*, *12*(133), 2-12.

- Kiss, C., Szilagyi, M., Paksy, A., & Poor, G. (2002). Risk factors for diffuse idiopathic skeletal hyperostosis: a case–control study. *Rheumatology*, *41*(1), 27-30.
- Klemensen, M. F. (2009). *Arkæologisk bygherrerapport: HOM 1649, Ole Worms Gade og Havneallé*. Retrieved from Horsens, DK:
- Klepinger, L. L., Katz, D., Micozzi, M. S., & Carroll, L. (1992). Evaluation of cast methods for estimating age from the os pubis. *Journal of Forensic Science*, *37*(3), 763-770.
- Kohn, R. R. (1978). (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kokich, V. G. (1976). Age changes in the human frontozygomatic suture from 20 to 95 years. *American journal of orthodontics*, 69(4), 411-430.
- Komar, D. A., & 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, L. W., & Frankenberg, S. R. (1992). Estimation of age structure in anthropological demography. *American Journal of Physical Anthropology*, 89(2), 235-256.
- Konigsberg, L. W., & Frankenberg, S. R. (1994). Paleodemography: "Not quite dead". *Evolutionary Anthropology: Issues, News, and Reviews, 3*(3), 92-105.
- Konigsberg, L. W., Herrmann, N. P., Wescott, D. J., & Kimmerle, E. H. (2008). Estimation and evidence in forensic anthropology: age-at-death. *Journal of Forensic Sciences*, 53(3), 541-557.
- Kunos, C. A., Simpson, S. W., Russell, K. F., & Hershkovitz, I. (1999). First rib metamorphosis: its possible utility for human age-at-death estimation. *American Journal of Physical Anthropology*, 110(3), 303-323.
- Kurki, H. (2005). Use of the first rib for adult age estimation: a test of one method. *International Journal of Osteoarchaeology*, *15*(5), 342-350.
- L'Abbé, E. N., Loots, M., & Meiring, J. H. (2005). The Pretoria bone collection: a modern South African skeletal sample. *HOMO-Journal of Comparative Human Biology*, *56*(2), 197-205.
- Lahdenperä, M., Lummaa, V., Helle, S., Tremblay, M., & Russell, A. F. (2004). Fitness benefits of prolonged post-reproductive lifespan in women. *Nature*, *428*(6979), 178-181.
- Lahdenperä, M., Russell, A. F., & Lummaa, V. (2007). Selection for long lifespan in men: benefits of grandfathering? *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1624), 2437-2444.
- Lahdenperä, M., Russell, A. F., Tremblay, M., & Lummaa, V. (2011). Selection on menopause in two premodern human populations: no evidence for the mother hypothesis. *Evolution*, *65*(2), 476-489.

- Larsen, C. S. (1997). Activity patterns 1. Articular and muscular modifications *Bioarchaeology: Interpreting behavior from the human skeleton* (pp. 161-194). New York, NY: Cambridge University Press.
- Leonetti, D. L., Nath, D. C., Heman, N. S., & Neill, D. B. (2005). Kinship Organization and the Impact of Grandmothers on Reproductive Success among the Matrilineal Khasi and Patrilineal Bengali of Northeast India. In E. Voland, A. Chasiotis, & W. Schiefenhovel (Eds.), *Grandmotherhood* (pp. 194–214). New Brunswick, NJ: Rutgers University Press.
- Lewis, M. E., & Flavel, A. (2006). Age Assessment of Child Skeletal Remains in Forensic Contexts. In A. Schmitt, E. Cunha, & J. Pinheiro (Eds.), *Forensic Anthropology and Medicine: Complementary Sciences from Recovery to Cause of Death* (pp. 243-257). Totowa, NJ: Humana Press Inc.
- Listi, G. A. (2016). The Use of Entheseal Changes in the Femur and Os Coxa for Age Assessment. *Journal of Forensic Sciences*, *61*(1), 12-18.
- Listi, G. A., & Manhein, M. H. (2012). The use of vertebral osteoarthritis and osteophytosis in age estimation. *Journal of Forensic Sciences*, *57*(6), 1537-1540.
- Livi-Bacci, M. (2012). A concise history of world population (5th ed.). Malden, MA: John Wiley and Sons.
- Loth, S. R. (1995). Age assessment of the Spitalfields cemetery population by rib phase analysis. *American Journal of Human Biology, 7*(4), 465-471.
- Loth, S. R., & İşcan, M. Y. (1989). Morphological assessment of age in the adult: the thoracic region. In M. Y. İşcan (Ed.), *Age markers in the human skeleton* (pp. 105-136). Springfield, IL: Charles C. Thomas.
- Loth, S. R., İşcan, M. Y., & Scheuerman, E. H. (1994). Intercostal variation at the sternal end of the rib. *Forensic science international*, *65*(2), 135-143.
- Lottering, N., MacGregor, D. M., Meredith, M., Alston, C. L., & Gregory, L. S. (2013). Evaluation of the Suchey–Brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. *American Journal of Physical Anthropology*, *150*(3), 386-399.
- Lovejoy, C. O., Meindl, R. S., Mensforth, R. P., & Barton, T. J. (1985). Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. *American Journal of Physical Anthropology, 68*(1), 1-14.
- Lovejoy, C. O., Meindl, R. S., Pryzbeck, T. R., Barton, T. S., Heiple, K. G., & Kotting, D. (1977). Paleodemography of the Libben site, Ottawa county, Ohio. *Science*, *198*(4314), 291-293.
- Lovejoy, C. O., Meindl, R. S., Pryzbeck, T. R., & Mensforth, R. P. (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(1), 15-28.

- Lynnerup, N., & Jacobsen, J. C. B. (2003). Brief communication: age and fractal dimensions of human sagittal and coronal sutures. *American Journal of Physical Anthropology, 121*(4), 332-336.
- Maat, G. (1987). Practising methods of age determination. Comments on methods combining multiple age indicators. *International Journal of Anthropology*, *2*(4), 293-299.
- Mahakkanukrauh, P., Khanpetch, P., Prasitwattanseree, S., & Case, D. T. (2013).

 Determination of sex from the proximal hand phalanges in a Thai population. *Forensic science international*, 226(1), 208-215.
- Mann, R. W., Jantz, R. L., Bass, W. M., & Willey, P. S. (1991). Maxillary suture obliteration: a visual method for estimating skeletal age. *Journal of Forensic Science*, *36*(3), 781-791.
- Mann, R. W., Symes, S. A., & Bass, W. M. (1987). Maxillary suture obliteration: aging the human skeleton based on intact or fragmentary maxilla. *Journal of Forensic Science*, 32(1), 148-157.
- Marlowe, F. (2000). The patriarch hypothesis. *Human Nature*, 11(1), 27-42.
- Martins, R., Oliveira, P. E., & Schmitt, A. (2012). Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. *Forensic science international*, 219(1), 287. e281-287.
- Martrille, L., Ubelaker, D. H., Cattaneo, C., Seguret, F., Tremblay, M., & Baccino, E. (2007). Comparison of four skeletal methods for the estimation of age at death on white and black adults. *Journal of Forensic Sciences*, *52*(2), 302-307.
- Masset, C. (1971). Erreurs systématiques dans la détermination de l'âge par les sutures crâniennes. *Bulletins et Mémoires de la Société d'anthropologie de Paris, 7*(1), 85-105.
- Masset, C. (1982). Estimation de l'age de deces par les sutures craniennes. (Thèse de Sciences Naturelles), Université de Paris VII. (Multigraphée)
- Masset, C. (1989). Age estimation on the basis of cranial sutures. In M. Y. Iscan (Ed.), *Age markers in the human skeleton* (pp. 71-103). Springfield, IL: Charles C. Thomas.
- Maxwell Museum of Anthropology Laboratory of Human Osteology. (2010). Documented Skeletal Collection. Retrieved from http://www.unm.edu/~osteolab/coll_doc.html
- Mays, S. (2015). The effect of factors other than age upon skeletal age indicators in the adult. *Annals of human biology, 42*(4), 332-341.
- McCormick, W. F. (1980). Mineralization of the costal cartilages as an indicator of age: preliminary observations. *Journal of Forensic Science*, *25*(4), 736-741.
- McCormick, W. F., & Stewart, J. H. (1988). Age related changes in the human plastron: a roentgenographic and morphologic study. *Journal of Forensic Science*, *33*(1), 100-120.
- McKern, T. W., & Stewart, T. D. (1957). Skeletal age changes in young American males analysed from the standpoint of age identification. Retrieved from

- Meindl, R. S., & Lovejoy, C. O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology*, *68*(1), 57-66.
- Meindl, R. S., & Lovejoy, C. O. (1989). Age changes in the pelvis: implications for paleodemography. In M. Y. İşcan (Ed.), *Age markers in the human skeleton* (pp. 137-168). Springfield, IL: Charles C. Thomas.
- Meindl, R. S., Lovejoy, C. O., Mensforth, R. P., & Walker, R. A. (1985). A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. *American Journal of Physical Anthropology*, 68(1), 29-45.
- Merritt, C. E. (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.
- Merritt, C. E. (2015). The influence of body size on adult skeletal age estimation methods. *American Journal of Physical Anthropology, 156*(1), 35-57.
- Michelson, N. (1934). The calcification of the first costal cartilage among whites and negroes. *Human biology*, 6(3), 543.
- Milella, M., Belcastro, M. G., Zollikofer, C. P. E., & Mariotti, V. (2012). The Effect of Age, Sex, and Physical Activity on Entheseal Morphology in a Contemporary Italian Skeletal Collection. *American Journal of Physical Anthropology*, *148*, 379-388.
- Milner, G. R., & Boldsen, J. L. (2012a). Estimating Age and Sex from the Skeleton, a Paleopathological. *A companion to paleopathology*, *34*, 268.
- Milner, G. R., & Boldsen, J. L. (2012b). Skeletal age estimation: where we are and where we should go. *A Companion to Forensic Anthropology*, 224-238.
- Milner, G. R., & Boldsen, J. L. (2012c). Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology, 148*(1), 98-110.
- Milner, G. R., Boldsen, J. L., Ousley, S. D., Getz, S. M., Weise, S., & Tarp, P. (2016). *Estimating Age from Adult Skeletons: New Directions in Transition Analysis Using a Wide Array of Traits*. Paper presented at the American Association of Physical Anthropologists, Atlanta, GA.
- Milner, G. R., Boldsen, J. L., Ousley, S. D., Weise, S., Getz, S. M., & Tarp, P. (2016). *Improved Adult Age Estimation Using New Skeletal Traits and Transition Analysis*. Paper presented at the American Academy of Forensic Sciences, Las Vegas, NV.
- Milner, G. R., Wood, J. W., & Boldsen, J. L. (2008). Advances in paleodemography. In M. A. Katzenberg & S. R. Saunders (Eds.), *Biological Anthropology of the Human Skeleton, Second Edition* (pp. 561-600): John Wiley & Sons, Inc.
- Molleson, T. (1995). Rates of ageing in the eighteenth century. *Grave Reflections: Portraying the Past Through Cemetery Studies*, 199-222.

- Moraitis, K., Zorba, E., Eliopoulos, C., & Fox, S. C. (2014). A test of the revised auricular surface aging method on a modern European population. *Journal of Forensic Sciences*, 59(1), 188-194.
- Moskovitch, G., Dedouit, F., Braga, J., Rougé, D., Rousseau, H., & Telmon, N. (2010). Multislice computed tomography of the first rib: a useful technique for bone age assessment. *Journal of Forensic Sciences*, *55*(4), 865-870.
- Mulhern, D. M., & Jones, E. B. (2005). Test of revised method of age estimation from the auricular surface of the ilium. *American Journal of Physical Anthropology*, *126*(1), 61-65.
- Mundy, G. R. (1999). Osteoporosis *Bone Remodeling and Its Disorders* (2nd ed., pp. 193-239). London,UK: Martin Dunitz Ltd.
- Murray, K. A., & Murray, T. (1991). A test of the auricular surface aging technique. *Journal of Forensic Science*, *36*(4), 1162-1169.
- N'Guyen, T., Gorse, F. C., & Vacher, C. (2007). Anatomical modifications of the mid palatal suture during ageing: a radiographic study. *Surgical and Radiologic Anatomy*, *29*(3), 253-259.
- Nagaoka, T., & Hirata, K. (2008). Demographic structure of skeletal populations in historic Japan: a new estimation of adult age-at-death distributions based on the auricular surface of the ilium. *Journal of archaeological science*, *35*(5), 1370-1377.
- Nagar, Y., & Hershkovitz, I. (2004). Interrelationship between various aging methods, and their relevance to palaeodemography. *Human evolution*, *19*(2), 145-155.
- National Institute of Justice. (2014). Research and Development in Forensic Science for Criminal Justice Purposes. (Grants.gov #: NIJ–2014-3744). Office of Justice Programs Retrieved from https://www.ncjrs.gov/pdffiles1/nij/sl001082.pdf.
- National Research Council. (2009). Strengthening forensic science in the United States: A path forward. Washington, DC: National Academies Press.
- Nawrocki, S. P. (1995). Taphonomic Processes in Historic Cemeteries. In A. L. Grauer (Ed.), Bodies of Evidence: Reconstructing History through Skeletal Analysis (pp. 49-66). New York, NY: Wiley-Liss.
- Nawrocki, S. P. (1998). Regression formulae for estimating age at death from cranial suture closure. Forensic osteology: advances in the identification of human remains. Charles C. Thomas, Springfield, 276-292.
- Neel, J. V., & Weiss, K. M. (1975). The genetic structure of a tribal population, the Yanomama Indians. XII. Biodemographic studies. *American Journal of Physical Anthropology, 42*(1), 25-51.
- Nikita, E. (2013). Quantitative assessment of the sternal rib end morphology and implications for its application in aging human remains. *Journal of Forensic Sciences*, *58*(2), 324-329.

- Obert, M., Schulte-Geers, C., Schilling, R. L., Harth, S., Kläver, M., Traupe, H., . . . Verhoff, M. A. (2010). High-resolution flat-panel volumetric CT images show no correlation between human age and sagittal suture obliteration—independent of sex. *Forensic science international*, 200(1), 180. e181-180. e112.
- Oettlé, A. C., & Steyn, M. (2000). Age estimation from sternal ends of ribs by phase analysis in South African blacks. *Journal of Forensic Science*, *45*(5), 1071-1079.
- Orrman, E. (2016). Growth and stagnation of population and settlement. In E. I. Kouri & J. E. Olesen (Eds.), *The Cambridge History of Scandinavia, Volume 2: 1520–1870* (Vol. 2, pp. 133–134): Cambridge University Press.
- Osborne, D. L., Simmons, T. L., & Nawrocki, S. P. (2004). Reconsidering the auricular surface as an indicator of age at death. *Journal of Forensic Science*, 49(5), JFS2003348-2003347.
- Ousley, S. D., & Jantz, R. L. (2012). Fordisc 3 and statistical methods for estimating sex and ancestry. *A Companion to Forensic Anthropology*, 311-329.
- Ousley, S. D., & Jantz, R. L. (2015). Fordisc 3.1 (Version 3.1.307). Retrieved from http://math.mercyhurst.edu/~sousley/Fordisc/
- Overbury, R. S., Cabo, L. L., Dirkmaat, D. C., & Symes, S. A. (2009). Asymmetry of the os pubis: Implications for the Suchey-Brooks method. *American Journal of Physical Anthropology*, 139(2), 261-268.
- Pal, G., & Tamankar, B. (1983). Preliminary study of age changes in Gujarati (Indian) pubic bones. *The Indian journal of medical research*, 78, 694.
- Parsons, F., & Box, C. (1905). The relation of the cranial sutures to age. *The journal of the Anthropological Institute of Great Britain and Ireland*, *35*, 30-38.
- Pasquier, E., De Saint Martin Pernot, L., Burdin, V., Mounayer, C., Le Rest, C., Colin, D., . . . Baccino, E. (1999). Determination of age at death: Assessment of an algorithm of age prediction using numerical three-dimensional CT data from pubic bones. *American Journal of Physical Anthropology*, 108(3), 261-268.
- Passalacqua, N. V. (2009). Forensic Age-at-Death Estimation from the Human Sacrum. *Journal of Forensic Sciences*, *54*(2), 255-262.
- Passalacqua, N. V. (2010). The Utility of the Samworth and Gowland Age-at-Death "Look-up" Tables in Forensic Anthropology. *Journal of Forensic Sciences*, *55*(2), 482-487.
- Pearson, O. M., & Buikstra, J. E. (2006). Behavior and the bones. In J. E. Buikstra & L. A. Beck (Eds.), *Bioarchaeology: the contextual analysis of human remains* (pp. 207-225). New York, NY: Elsevier.
- Peccei, J. S. (1995). The origin and evolution of menopause: the altriciality-lifespan hypothesis. *Ethology and Sociobiology, 16*(5), 425-449.

- Peccei, J. S. (2001a). A critique of the grandmother hypotheses: old and new. *American Journal of Human Biology*, *13*(4), 434-452.
- Peccei, J. S. (2001b). Menopause: adaptation or epiphenomenon? *Evolutionary Anthropology: Issues, News, and Reviews, 10*(2), 43-57.
- Peccei, J. S. (2005). Menopause: Adaption and Epiphenomenon. In E. Voland, A. Chasiotis, & W. Schiefenhovel (Eds.), *Grandmotherhood* (pp. 38-58). New Brunswick, NJ: Rutgers University Press.
- Pedersen, D. D., & Boldsen, J. (2010). *Antropologisk Rapport: HOM1649, Ole Worms Gade*. Retrieved from Odense. DK:
- Perizonius, W. (1984). Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (AD 1883–1909). *Journal of Human Evolution, 13*(2), 201-216.
- Petersen, C. G. (2007). Bygherrerapport I. Undersøgelse af: Klosterkom-pleks og efterreformatoriske grave. Retrieved from
- Petersen, W., Braidwood, R. J., Dobyns, H. F., Eberhard, W., Kennedy, R. E., Kurth, G., . . . Wilkinson, P. F. (1975). A Demographer's View of Prehistoric Demography [and Comments and Replies]. *Current Anthropology*, *16*(2), 227-245.
- Pfeiffer, S. (1980). Bone-Remodelling Age Estimates Compared With Estimates by Other Techniques. *Current Anthropology*, *21*(6), 793-794.
- Pfeiffer, S. (1985). Comparison of adult age estimation techniques, using an ossuary sample. Canadian Journal of Anthropology, 4(12), 13-17.
- Poelstra, K. A. (2013). Diffuse Idiopathic Skeletal Hyperostosis. In D. G. Anderson & A. R. Vaccaro (Eds.), *Decision making in spinal care* (2nd ed., pp. 191-195). New York, NY: Thieme.
- Powers, R. (1962). The Disparity Between Known Age and Age as Estimated by Cranial Suture Closure. *Man*, 62, 52-54.
- Purves, S., Woodley, L., & Hackman, L. (2011). Age determination in the adult *Forensic Anthropology*–2000 to 2010 (pp. 29-59): CRC Press Boca Raton, FL.
- Rejtarová, O., Hejna, P., Soukup, T., & Kuchař, M. (2009). Age and sexually dimorphic changes in costal cartilages. A preliminary microscopic study. *Forensic science international*, 193(1), 72-78.
- Resnick, D., Niwayama, G., & Goergen, T. G. (1975). Degenerative disease of the sacroiliac joint. *Investigative radiology*, *10*(6), 608-621.
- Riebel, F. (1929). Ossification of the costal cartilages: their relation to habitus and disease. *Am J Roentgenol Radium Ther*, 21, 44-47.
- Rissech, C., Estabrook, G. F., Cunha, E., & Malgosa, A. (2006). Using the acetabulum to estimate age at death of adult males. *Journal of Forensic Sciences*, *51*(2), 213-229.

- Rissech, C., Estabrook, G. F., Cunha, E., & Malgosa, A. (2007). Estimation of Age-at-Death for Adult Males Using the Acetabulum, Applied to Four Western European Populations. *Journal of Forensic Sciences*, *52*(4), 774-778.
- Rissech, C., Wilson, J., Winburn, A. P., Turbón, D., & 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.
- Ross, S. M. (2010). Random Variables *A First Course in Probability* (8th ed., pp. 117-185). Supper Saddle River, NJ: Pearson.
- Rougé-Maillart, C., Telmon, N., Rissech, C., Malgosa, A., & Rougé, D. (2004). The determination of male adult age at death by central and posterior coxal analysis—a preliminary study. *Journal of Forensic Science*, 49(2), 1-7.
- Rougé-Maillart, C., Vielle, B., Jousset, N., Chappard, D., Telmon, N., & Cunha, E. (2009). Development of a method to estimate skeletal age at death in adults using the acetabulum and the auricular surface on a Portuguese population. *Forensic science international*, 188(1), 91-95.
- Russell, K. F., Simpson, S. W., Genovese, J., Kinkel, M. D., Meindl, R. S., & Lovejoy, C. O. (1993). Independent test of the fourth rib aging technique. *American Journal of Physical Anthropology*, *92*(1), 53-62.
- Sahni, D., & Jit, I. (2005). Time of closure of cranial sutures in northwest Indian adults. *Forensic science international*, 148(2), 199-205.
- Saito, K., Shimizu, Y., & Ooya, K. (2002). Age-related morphological changes in squamous and parietomastoid sutures of human cranium. *Cells Tissues Organs*, *170*(4), 266-273.
- Sakaue, K. (2006). Application of the Suchey-Brooks system of pubic age estimation to recent Japanese skeletal material. *Anthropological science*, *114*(1), 59-64.
- Salem, N. H., Aissaoui, A., Mesrati, M. A., Belhadj, M., Quatrehomme, G., & Chadly, A. (2014). Age estimation from the sternal end of the fourth rib: A study of the validity of İşcan's Method in Tunisian male population. *Legal Medicine*, *16*(6), 385-389.
- Samworth, R., & Gowland, R. (2007). Estimation of adult skeletal age-at-death: statistical assumptions and applications. *International Journal of Osteoarchaeology, 17*(2), 174-188.
- San Millán, M., Rissech, C., & 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.
- Sashin, D. (1930). A critical analysis of the anatomy and the pathologic changes of the sacroiliac joints. *J Bone Joint Surg Am*, *12*(4), 891-910.

- Saunders, S., Fitzgerald, C., Rogers, T., Dudar, C., & McKillop, H. (1992). A test of several methods of skeletal age estimation using a documented archaeological sample. *Canadian Society of Forensic Science Journal*, 25(2), 97-118.
- Sauvage, H. E. (1870). Note sur l'etat senile du Crane. Bulletins de la Société d'anthropologie de Paris, 5(1), 576-586.
- Scheuer, J., & Bowman, J. (1995). Correlation of documentary and skeletal evidence in the St Brides crypt population. *Grave Reflections Portraying the Past through Cemetery Studies Canadian Scholar's Press Inc Toronto*, 49-70.
- 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), 1-6.
- Schmitt, A. (2005). Une nouvelle méthode pour estimer l'âge au décès des adultes à partir de la surface sacro-pelvienne iliaque. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*(17 (1-2)), 89-101.
- Schmitt, A. (2008). Une nouvelle méthode pour discriminer les individus décédés avant ou après 40 ans à partir de la symphyse publienne. *Journal de médecine légale droit médical*, 51(1), 15-24.
- Schmitt, A., & Broqua, C. (2000). Approche probabiliste pour estimer l'âge au décès à partir de la surface auriculaire de l'ilium. *Bulletins et Mémoires de la Société d'anthropologie de Paris*, 12(3-4), 279-301.
- Schmitt, A., & Murail, P. (2004). Is the first rib a reliable indicator of age at death assessment? Test of the method developed by Kunos et al (1999). *HOMO-Journal of Comparative Human Biology*, *54*(3), 207-214.
- Schmitt, A., Murail, P., Cunha, E., & Rougé, D. (2002). Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age at death estimation. *Journal of Forensic Science*, *47*(6), 1203-1209.
- Schmitt, A., Wapler, U., Couallier, V., & Cunha, E. (2007). Are bone losers distinguishable from bone formers in a skeletal series? Implications for adult age at death assessment methods. *HOMO-Journal of Comparative Human Biology, 58*(1), 53-66.
- Schmitt, H., & Tamáska, L. (1970). Beiträge zur forensischen Osteologie. Zeitschrift für Rechtsmedizin, 67(4), 230-248.
- Scoles, P. V., Salvagno, R., Villalba, K., & Riew, D. (1987). Relationship of iliac crest maturation to skeletal and chronologic age. *Journal of pediatric orthopedics*, *8*(6), 639-644.
- Seeman, E. (2008). Modeling and Remodeling: The Cellular Machinery Responsible for the Gain and Loss of Bone's Material and Structural Strength. In J. P. Bilezikian, L. G. Raisz, & T. J. Martin (Eds.), *Principles of Bone Biology* (3rd ed., pp. 3-28). New York, NY: Academic Press.

- Semine, A. A., & Damon, A. (1975). Costochondral ossification and aging in five populations. *Human biology*, 101-116.
- Shanley, D. P., & Kirkwood, T. B. (2001). Evolution of the human menopause. *Bioessays*, 23(3), 282-287.
- Shanley, D. P., Sear, R., Mace, R., & Kirkwood, T. B. (2007). Testing evolutionary theories of menopause. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1628), 2943-2949.
- Sharma, G., Gargi, J., Kalsey, G., Singh, D., Rai, H., & Sandhu, R. (2008). Determination of age from pubic symphysis: an autopsy study. *Medicine, Science and the Law, 48*(2), 163-169.
- Shirley, N. R., & Ramirez Montes, P. A. (2015). Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component-Based Methods. *Journal of Forensic Sciences*, *60*(1), 107-111.
- Simpson, G. G. (1949). *The Meaning of Evolution*. New Haven, CT: Yale University Press.
- Singer, R. (1953). Estimation of age from cranial suture closure. A report on its unreliability. *Journal of Forensic Medicine*, 1(1), 52-59.
- Singh, P., Oberoi, S., Gorea, R., & Kapila, A. (2004). Age estimation in old individuals by CT scan of skull. *Journal of Indian Academy of Forensic Medicine*, *26*(1), 10-13.
- Sinha, A., & Gupta, V. (1995). A study on estimation of age from pubic symphysis. *Forensic science international*, 75(1), 73-78.
- Sitchon, M. L., & Hoppa, R. D. (2005). Assessing age-related morphology of the pubic symphysis from digital images versus direct observation. *Journal of Forensic Science*, 50(4), JFS2004182-2004185.
- Sjvold, T. (1975). Tables of the combined method for determination of age at death given by Nemeskri, Harsnyi, Acsdi. *Anthrop Kzl, 19*, 922.
- Skjærvø, G. R., & Røskaft, E. (2013). Menopause: no support for an evolutionary explanation among historical Norwegians. *Experimental Gerontology, 48*(4), 408-413.
- Slice, D. E., & Algee-Hewitt, B. F. (2015). Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. *Journal of Forensic Sciences*, 60(4), 835-843.
- Snodgrass, J. J. (2004). Sex differences and aging of the vertebral column. *Journal of Forensic Science*, 49(3), JFS2003198-2003196.
- Snow, C. C. (1983). Equations for estimating age at death from the pubic symphysis: a modification of the McKern-Stewart method. *Journal of Forensic Science*, *28*(4), 864-870.

- Sogner, S. (2016). Demography and family,c.1650–1815. In E. I. Kouri & J. E. Olesen (Eds.), The Cambridge History of Scandinavia, Volume 2: 1520–1870 (Vol. 2, pp. 439–440): Cambridge University Press.
- St. Bride's Church. (2016). St. Bride's:History. Retrieved from http://www.stbrides.com/history/
- Stewart, J. H., & McCormick, W. F. (1984). A sex-and age-limited ossification pattern in human costal cartilages. *American journal of clinical pathology*, *81*(6), 765-769.
- Stewart, T. D. (1957). Distortion of the pubic symphyseal surface in females and its effect on age determination. *American Journal of Physical Anthropology*, *15*(1), 9-18.
- Stewart, T. D. (1958). Rate of development of vertebral osteoarthritis in American whites and its significance in skeletal age identification.
- Storey, R. (2007). An elusive paleodemography? A comparison of two methods for estimating the adult age distribution of deaths at late Classic Copan, Honduras. *American Journal of Physical Anthropology*, 132(1), 40-47.
- Stoyanova, D., Algee-Hewitt, B. F., & Slice, D. E. (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*(3), 431-440.
- Suchey, J. M. (1979). Problems in the aging of females using the os pubis. *American Journal of Physical Anthropology*, *51*(3), 467-470.
- Suchey, J. M., Brooks, S. T., & Katz, D. (1988). Instructions for use of the Suchey-Brooks system for age determination of the female os pubis *Instructional materials* accompanying female pubic symphysial models of the Suchey-Brooks system. Fort Collins, Colorado: France Casting.
- Suchey, J. M., Wiseley, D. V., & Katz, D. (1986). Evaluation of the Todd and McKern-Stewart methods for aging the male os pubis. *Forensic osteology. Advances in the identification of human remains, 1st edn. Charles C Thomas, Springfield.*
- Susser, M. (1985). Epidemiology in the United States after World War II: the evolution of technique: Columbia University.
- Tarp, P., & Boldsen, J. L. (2010). *Antropologisk Rapport: HOM1272 Horsens Klosterkirke*. Retrieved from Odense, DK:
- Team, R. C. D. (2008). R: A language and environment for statistical computing (Version 2.15.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org
- Telmon, N., Gaston, A., Chemla, P., Blanc, A., Joffre, F., & Rougé, D. (2005). Application of the Suchey-Brooks method to three-dimensional imaging of the pubic symphysis. *Journal of Forensic Science*, *50*(3), 1-6.

- Telmon, N., Rougé, D., Brugne, J., Pujol, J., Larrouy, G., & Arbus, L. (1996). Comparison of a bone criterion and a dental criterion for estimation of age at death in the study of an ancient cemetery. *International Journal of Anthropology, 11*(1), 1-9.
- Todd, T. W. (1920). Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology*, *3*(3), 285-334.
- Todd, T. W. (1921). Age changes in the pubic bone. *American Journal of Physical Anthropology*, *4*(1), 1-70.
- Todd, T. W. (1923). Age changes in the pubic symphysis: VII. The anthropoid strain in human pubic symphyses of the third decade. *Journal of anatomy*, *57*(Pt 3), 274.
- Todd, T. W., & Lyon, D. (1924). Endocranial suture closure. Its progress and age relationship. Part I.—Adult males of white stock. *American Journal of Physical Anthropology, 7*(3), 325-384.
- Todd, T. W., & Lyon, D. (1925a). Cranial suture closure. Its progress and age relationship. Part II.—Ectocranial closure in adult males of white stock. *American Journal of Physical Anthropology*, 8(1), 23-45.
- Todd, T. W., & Lyon, D. (1925b). Cranial suture closure. Its progress and age relationship. Part III.—Endocranial closure in adult males of Negro stock. *American Journal of Physical Anthropology*, 8(1), 47-71.
- Townsend, N., & Hammel, E. (1990). Age estimation from the number of teeth erupted in young children: an aid to demographic surveys. *Demography*, *27*(1), 165-174.
- Tuljapurkar, S. D., Puleston, C. O., & Gurven, M. D. (2007). Why men matter: mating patterns drive evolution of human lifespan. *PloS one*, *2*(8), e785.
- Ubelaker, D. (1995). Historic Cemetery Analysis: Practical Considerations. In A. L. Grauer (Ed.), Bodies of Evidence: Reconstructing History through Skeletal Analysis (pp. 37-48). New York, NY: Wiley-Liss.
- Ubelaker, D. H. (1995). Historic Cemetery Analysis: Practical Considerations. In A. L. Grauer (Ed.), *Bodies of Evidence: Reconstructing History through Skeletal Analysis* (pp. 37-48). New York, NY: Wiley-Liss.
- Ubelaker, D. H. (2000). Methodological considerations in the forensic applications of human skeletal biology. *Biological Anthropology of the Human Skeleton. Wiley-Liss, New York, NY*, 41-67.
- Ubelaker, D. H. (2008). Issues in the global applications of methodology in forensic anthropology. *Journal of Forensic Sciences*, *53*(3), 606-607.
- University of Iowa, & Office of the State Archaeologist. (2012a). UI-Stanford Collection Management Policy (pp. 1-3). Iowa City, IA.
- University of Iowa, & Office of the State Archaeologist. (2012b). The University of Iowa-Stanford Collection FAQ Sheet (pp. 1). Iowa City, IA.

- Usher, B. M. (2002). Reference samples: the first step in linking biology and age in the human skeleton. In R. Hoppa & J. W. Vaupel (Eds.), *Paleodemography: age distribution from skeletal samples* (pp. 29-47). Cambridge: Cambridge University Press.
- Vahtola, J. (2003). Population and settlement. In K. Helle (Ed.), *The Cambridge History of Scandinavia* (Vol. 1, pp. 559-580). Cambridge: Cambridge University Press.
- Van der Merwe, A., Işcan, M., & L'Abbè, E. (2006). The pattern of vertebral osteophyte development in a South African population. *International Journal of Osteoarchaeology*, 16(5), 459-464.
- Van Gerven, D. P., & Armelagos, G. J. (1983). "Farewell to paleodemography?" Rumors of its death have been greatly exaggerated. *Journal of Human Evolution*, *12*(4), 353-360.
- Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty On the Dynamics of Mortality. *Demography*, *16*(3), 439-454.
- Verzeletti, A., Cassina, M., Micheli, L., Conti, A., & De Ferrari, F. (2010). Age estimation from the rib by components method analysis in white males. *The American journal of forensic medicine and pathology, 31*(1), 27-33.
- Verzeletti, A., Terlisio, M., & De Ferrari, F. (2013). Age-at-death estimation in Caucasian females from the morphological analysis of the sternal end of the fourth rib. *Legal Medicine*, *15*(1), 47-49.
- Villa, C., Buckberry, J., Cattaneo, C., & Lynnerup, N. (2013). Technical Note: Reliability of Suchey-Brooks and Buckberry-Chamberlain methods on 3D visualizations from CT and laser scans. *American Journal of Physical Anthropology, 151*(1), 158-163.
- Walker, P. L. (1995). Problems of Preservation and Sexism in Sexing: Some Lessons From Historical Collections for Paleodemographers. In S. R. Saunders & A. Herring (Eds.), *Grave Reflections: Portraying the Past Through Cemetery Studies* (pp. 31-47). Toronto, CA: Canadian Scholars' Press.
- Wang, F.-I., Wang, G., & Tian, X.-m. (2012). Progress in the study of estimation of age from pubic symphsis [J]. *Forensic Science and Technology, 3*, 020.
- Wärmländer, S. K., & Sholts, S. B. (2011). Sampling and statistical considerations for the Suchey–Brooks method for pubic bone age estimation: Implications for regional comparisons. *Science & Justice*, *51*(3), 131-134.
- Warner, D. C. (2012). Access to health services for immigrants in the USA: from the Great Society to the 2010 Health Reform Act and after. *Ethnic and Racial Studies*, *35*(1), 40-55.
- Weber, J. (2012). The evolving interstate highway system and the changing geography of the United States. *Journal of Transport Geography*, 25, 70-86.

- Weise, S., Boldsen, J. L., Gampe, J., & Milner, G. R. (2012). How old was 'Geriatrix'? Estimating the age of old individuals with Calibrated Expert Inference. Paper presented at the 81st Annual Meeting of the American Association of Physical Anthropologists, Portland, Oregon.
- Weismann, A. (1891). The duration of life. In E. B. Poulton, S. Schönland, & A. E. Shipley (Eds.), *Essays Upon Heredity and Kindred Biological Problems* (Second ed., Vol. 1, pp. 1-66). Oxford, UK: Clarendon Press.
- Weiss, K. M. (1981). Evolutionary perspectives on human aging. In P. T. Amoss & S. Harrell (Eds.), *Other Ways of Growing Old* (pp. 25-25l). Stanford, CA: Stanford University Press.
- Welch, R., & Mitchell, P. (2000). Food processing: a century of change. *British medical bulletin*, 56(1), 1-17.
- Wenguang, M. S. Z., & Ke, L. S. W. X. Q. (1989). The Relation Between the Age and Changes in Chinese Skull Suture *Acta Anthropologica Sinica*, *2*, 005.
- Wescott, D. J., & Drew, J. L. (2015). Effect of obesity on the reliability of age-at-death indicators of the pelvis. *American Journal of Physical Anthropology*, *156*(4), 595-605.
- Williams, G. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolutzon 11: 398-411. 1966 Adaptation and Natural Selection: Princeton, New Jersey: Princeton University Press.
- Wink, A. E. (2014). Pubic symphyseal age estimation from three-dimensional reconstructions of pelvic CT scans of live individuals. *Journal of Forensic Sciences*, *59*(3), 696-702.
- Wittschieber, D., Schulz, R., Vieth, V., Küppers, M., Bajanowski, T., Ramsthaler, F., . . . Schmeling, A. (2014). The value of sub-stages and thin slices for the assessment of the medial clavicular epiphysis: a prospective multi-center CT study. *Forensic science, medicine, and pathology, 10*(2), 163-169.
- Wolff, K., Vas, Z., Sótonyi, P., & Magyar, L. (2012). Skeletal age estimation in Hungarian population of known age and sex. *Forensic science international, 223*(1), 374. e371-374. e378.
- Wood, J. W., Milner, G. R., Harpending, H. C., Weiss, K. M., Cohen, M. N., Eisenberg, L. E., . . . Wilkinson, R. G. (1992). The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples [and Comments and Reply]. *Current Anthropology, 33*(4), 343-370.
- Wood, J. W., & Smouse, P. E. (1982). A method of analyzing density-dependent vital rates with an application to the Gainj of Papua New Guinea. *American Journal of Physical Anthropology*, *58*(4), 403-411.
- Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673-686. doi:10.1198/01621450400000980

- Wood, S. N. (2006a). GAMS in Practice: mgcv *Generalized Additive Models: An Introduction with R* (pp. 221-275). Boca Raton, FL: Chapman & Hall/CRC.
- Wood, S. N. (2006b). *Generalized Additive Models: An Introduction with R*. New York, NY: Chapman & Hall/CRC.
- Wood, S. N. (2006c). Generalized Linear Models *Generalized Additive Models: An Introduction with R* (pp. 59-120). New York, NY: Chapman & Hall/CRC.
- Wood, S. N. (2006d). Introducing GAMs *Generalized Additive Models: An Introduction with R* (pp. 121-143). New York, NY: Chapman & Hall/CRC.
- Wood, S. N. (2006e). Linear Models *Generalized Linear Models: An Introduction with R* (pp. 1-55). New York, NY: Chapman & Hall/CRC.
- Yavuz, M. F., İşcan, M. Y., & Çöloğlu, A. S. (1998). Age assessment by rib phase analysis in Turks. *Forensic science international*, *98*(1), 47-54.
- Yoder, C., Ubelaker, D. H., & Powell, J. F. (2001). Examination of variation in sternal rib end morphology relevant to age assessment. *Journal of Forensic Science*, *46*(2), 223-227.
- Zhaojin, L. W. C. S. X. (1988). Estimation of Age From The Pubic Symphysis of Chinese Male by Means of Multiple Analysis *Acta Anthropologica Sinica*, *2*, 009.
- Zhongyao, Z. (1986). A Further Study on the Relationship Between Morphologic Features of Pubic Symphyes and Age Estimation. *Acta Anthropologica Sinica*, *2*, 003.
- Zhongyao, Z. (1995). Study on Relation Between Variation of Male Pubic Micro-sturcutre and Age CHINESE JOURNAL OF FORENSIC MEDICINE, 4, 006.
- Zhongyao, Z., & Yang, G. (1999). Improvement of Equation for Age Estimation by Pubic Bone *CHINESE JOURNAL OF FORENSIC MEDICINE*, 3, 009.
- Zongyao, Z. (1982). A Preliminary Study of Estimation of Age by Morphological Changes in the Symphysis Pubis *Acta Anthropologica Sinica*, *2*, 004.

VITA Sara Marie Getz

EDUCATION

2017	PhD	Anthropology	The Pennsylvania State University
2011	MS	Biological & Forensic Anthropology	Mercyhurst University
2008	BS	Biological Anthropology	The Pennsylvania State University:
			Schreyer Honors College

PROFESSIONAL EXPERIENCE

2014-2016	Research Assistant	Penn State Dept. of Anthropology (GR Milner)			
2014	Visiting Scientist	New York City Office of Chief Medical Examiner			
2012-2014	Online Course Instructor	Penn State Dept. of Anthropology			
2011-2014	Graduate Assistant	Penn State Dept. of Anthropology			
2009-2011	Graduate Assistant	Mercyhurst Dept. of Applied Forensic Sciences			
2010	Museum Aide	Smithsonian NMNH & MSC			
2007	Intern	C.A. Pound Human Identification Lab			

PUBLICATIONS

- 2016. Milner GR, and **SM Getz**. "Skeletal Age Estimation." In Oxford Bibliographies in Anthropology. John Jackson Jr., editor. New York: Oxford University Press.
- 2012. Dirkmaat DC, GO Olsen, AR Klales, and **SM Getz**. The Role of Forensic Anthropology in the Recovery and Interpretation of the Fatal-Fire Victim. In: Dennis C. Dirkmaat, editor. A Companion to Forensic Anthropology. Malden, MA: Wiley-Blackwell p.113-135.

ABSTRACTS & CONFERENCE POSTERS

- 2016. Milner GR, JL Boldsen, SD Ousley, **SM Getz**, S Weise, and P Tarp. Estimating Age from Adult Skeletons: New Directions in Transition Analysis Using a Wide Array of Traits. American Journal of Physical Anthropology Supplement 159 (S62): 231.
- 2016. Milner GR, JL Boldsen, SD Ousley, S Weise, **SM Getz,** and P Tarp. Improved Adult Age Estimation Using New Skeletal Traits and Transition Analysis. Proceedings of the American Academy of Forensic Sciences 22: 57.
- 2015. **Getz SM**, GR Milner, and JL Boldsen. Skeletal age estimation for adults: Long-standing problems, recent developments, and a solution. American Journal of Physical Anthropology Supplement 156 (S60): 144.
- 2014. **Getz SM**, GR Milner, and JL Boldsen. Deconstructing Expert Judgment: Experience-Based Age Estimates and Their Role in Improving Transition Analysis (ADBOU). Society for Human Biology Symposium on Age Estimation. Oxford, UK. December 2014.
- 2011. **Getz SM**. Investigation and Critique of the DiGangi et al. (2009) First Rib Age-at-Death Estimation Method. Proceedings of the American Academy of Forensic Sciences 17: 350.

GRANTS & AWARDS

2016	\$400	Penn State Anthropology Dept. Conference Travel Grant
2015	\$23950	National Science Foundation Doctoral Dissertation Research Improvement
		Grant (NSF DDRIG) – Award #: 1455810 (GR Milner, advisor)
2014	\$600	Penn State Anthropology Dept. International Conference Travel Grant
2013	\$7412	Penn State Anthropology Dept. Post-Comprehensive Hill Fellowship Grant
2012	\$1053	Penn State Anthropology Dept. Pre-Comprehensive Hill Fellowship Grant
2011		Mercyhurst College Sister Eustace Taylor Graduate Student Award
2009	\$400	Mercyhurst College Faculty-Student Research Grant
2008	\$350	Schreyer Honors College Undergraduate Thesis Research Grant

PROFESSIONAL ORGANIZATION MEMBERSHIP

2014+	Student Member	American Anthropological Association (AAA)
2011+	Student Affiliate	American Academy of Forensic Sciences (AAFS)
2010+	Student Member	American Association of Physical Anthropologists (AAPA)