

**Understanding Population-Specific Age Estimation
Using Multivariate Cumulative Probit Regression
For Asian Skeletal Samples**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Ji Eun Kim
December 2016

Copyright © 2016 by Ji Eun Kim
All rights reserved.

Acknowledgements

I could not have been able to come this far without the help of numerous people. First, I cannot thank enough Dr. Dawnie W. Steadman, who has given me endless support, encouragement, and advice both in and outside the academia for almost 12 years. Thank you and thank you again, Dr. Steadman, for all your patience, all your efforts to instill more confidence in me, and most importantly, for showing me I can indeed do “something,” like finishing a Ph.D. I hope we can have another round of great adventures in Korea together in the near future!

I would also like to thank Dr. Lee M. Jantz for all her extremely thoughtful guidance at the Forensic Anthropology Center (FAC). In 2011, when I first joined the FAC as a first year Ph.D. student, I was very nervous and hoped I could contribute to the extensive FAC operations. I would not have been able to accomplish as much as I did at the FAC without a help of Lee. Working with her has been not only truly enlightening, but also remarkably delightful. I utterly admire her 20+ years of knowledge on the FAC body donation program, and also her brilliant sense of humor. You are indeed the muse of the FAC!

I would take this opportunity to thank Dr. Graciela S. Cabana for introducing me to the study of population genetics that entirely changed my perspective on the concepts of race, ethnicity and ancestry, and also my identity as an international living in the U.S. I must admit that population genetics was the most difficult subject among all other topics for my preliminary exams, but it was definitely worth the effort.

I would also like to express my sincere appreciation to Dr. James A. Fordyce for spending, by default, two-hour advising sessions concerning my statistical concerns. Thank you for all your flexibility and availability every time I was desperately in need of your advice, and

for providing me exceptionally helpful suggestions that sometimes involved a whiteboard to make sure that I understood everything.

The fieldwork part of this project was entirely supported by several funding sources. I am particularly grateful to the Wenner-Gren Foundation Doctoral Dissertation Fieldwork Grant, the W. K. McClure Scholarship for the Study of World Affairs, University of Tennessee, and the W. M. Bass Endowment and the W. Leitner Award offered by Forensic Anthropology Center, University of Tennessee for enabling me to travel to Japan and Thailand to complete my data collection.

This research would also not have been possible without the support of the professors and staff members of Japanese and Thai institutions housing the four skeletal collections used in this study. Especially, I thank Dr. Yoshiharu Matsuno at Chiba University, Chiba, Japan, not only for granting me access to the Chiba Collection, but also for his hospitality, for all the delicious lunch places he took me, and for the wonderful day trip to St. Marianna University.

I also deeply appreciate Dr. Yoshinori Kawai at Jikei University, Tokyo, Japan, and the most hardworking staff members at his lab. Even though I was always visiting Jikei at their busiest time of the year, they were always welcoming and willing to accommodate my schedule (even on Christmas!). Thank you very much, Murahata-san, Kato-san, Gucci, and Dr. Negishi for all your help in the collection.

I would also like to send a special thank you to my friend, Ms. Atsuko Hayashi. My data collection in Japan would have been impossible without her help. She not only helped me create amazing academic connections in Japan, but also took good, motherly care of me while we were traveling together in Japan. I would never be able to forget the delicious Christmas Eve dinner at a tiny Peruvian restaurant that we randomly found at the corner of a busy street in Shinbashi!

I owe many thanks and great appreciation to Dr. Pasuk Mahakkanukrauh at Chiang Mai University, Chiang Mai, Thailand, and Dr. Panya Tuamsuk at Khon Kaen University (KKU), Khon Kaen, Thailand, for their generosity of granting me access to their wonderful collections. Dr. Pasuk and Dr. Panya always made sure that I was also enjoying cultural experiences while I was in Thailand. I also appreciate Dr. Panya for providing me many opportunities to make new friends at KKU, and for inviting me to many departmental events so that I did not feel isolated during my long research period in Khon Kaen. Thanks to Dr. Pasuk, Neu and Doi I was able to visit beautiful temples and mountains in Chiang Mai, and also was able to have the world's best Khao Soi noodles. I made so many unforgettable memories in Khon Kaen thanks to my wonderful friends, P'Nuch, Sam, Ai, Oi, and Arm. My trip to Thailand was better because of them.

Perhaps, the best part of my graduate school experience is that I have built such marvelous friendships that cannot be bought anywhere. I'm sending special thanks to Dr. Christina Fojas, and her awesome husband, Adam, my best Italian sister, Dr. Giovanna Vidoli and her super Mike and baby Matilda, Dr. Dr. (not a typo) Heli Maijanen, Dr. Beatrix Dudzik, Dr. Sean Tallman, Jake Smith, Dr. Audrey Scott, Dr. Janna Andronowski, Dr. Kyle McCormick, Katie Kulhavy, Dr. Mike Kenyherztz, and other great friends and colleagues in- and outside of UTK. I would not have been able to make it through the long marathon of a graduate program without them. Thank you for all your jokes and commiserations, and our numerous day- and night-outs, international dinner nights, Thanksgiving dinners, birthday bashes, and barbecues.

I also thank the kindest people on earth, my family, back in Korea. I thank my father, mother, and big brother (plus Noah!), who always give me infinitely unconditional love and non-exhausting support, cheers and happy thoughts all the way from the opposite side of the globe. I

would not have been able to overcome homesickness if they did not send me thousands of pictures and videos of my 8-year-old Jindo, Noah back home. The countless ups and downs of my life away from home would have been more difficult without their support, love and words of wisdom.

Lastly, I thank my soul mate, Yuki, for letting me experience what it is like to live in the moment, not the past or the future. I had completely forgotten this during the graduate school years. Thanks to him, now I know that hiking, lying on the beach, wandering around downtown, and spending hours to cook a perfect top-sirloin steak are not something to feel guilty about or slacking from work, but they are the secret ingredients for the perfect recipe to a happy, real life.

Abstract

For many years, the field of anthropology has encouraged anthropologists to assume that population variation exists in skeletal aging although interpretations of population specificity in skeletal aging have been inconsistent. This project investigates age progressive changes in modern East and Southeast Asian populations, and attempts to quantify the magnitude of differences or similarities in skeletal aging between different Asian groups as a first step to develop a more inclusive age estimation method for Asian populations. Specifically, this study explores the utility of currently available age estimation methods for Asian populations, asks whether a population-specific aging method should be region-specific (Thai vs. Japan) or continental-specific (Asian), and investigates whether population specificity in age estimation is needed at all.

To achieve these goals, data from four skeletal collections representing 20th century Japanese and modern Thais were collected. The four age estimation methods were applied: Transition Analysis (TA) (Boldsen et al. 2002b), plus three ‘conventional’ methods of Suchey and Katz (1998), Lovejoy et al. (1985a), and Meindl and Lovejoy (1985). To develop age estimation models for the Asian samples, while minimizing methodological error a multivariate ordered probit regression model was fitted to the Asian skeletal data under a Bayesian framework.

The results of this study show that Japanese-specific age estimation models do not necessarily increase error in age estimates of Thai populations, and vice versa. Thus, multivariate ordered probit models fitted on a pooled sample of Japanese and Thais yielded significantly improved age estimates. Although TA and conventional methods yielded reasonable age estimates for Japanese and Thai populations, the probit models of pooled Asian groups

outperformed those methods. The results are even more promising when the scores of TA and conventional methods were combined to fit a multivariate ordered probit model.

Broader implications of this research is that, at least for Japanese and Thais, a single age estimation method can be developed. In addition, any error and bias in ages at death estimation may not be due to differences in population-specificity in skeletal aging and age estimation error induced by biased reference samples may be greater than expected.

Table of Contents

Chapter 1 Introduction.....	1
Current study	6
Chapter organization.....	8
Chapter 2 The Evolution of Age Estimation Methods	10
Steps for developing an age-at-death estimation method.....	11
Early interest in age-related osteometamorphosis.....	14
Pubic symphysis.....	15
Sacroiliac joint	19
Cranial sutures	22
Validation studies.....	24
Problematic osteological techniques of conventional age estimation methods.....	26
Statistical problems with conventional aging methods.....	28
Influence of reference collections on the age-at-death distribution.....	28
Traditional statistical approaches and their assumptions.....	32
A possible solution: Bayes' Rule/Theorem	37
Transition Analysis	40
Chapter 3 The Population History of Southeast and East Asia.....	45
Why study Asians?.....	46
The peopling of Southeast (SE) and East (E) Asia.....	47
Why is knowing population history important for developing aging methods?	52
Hypotheses	57
Chapter 4 Methods	62

Skeletal samples	62
Data collection	63
Scoring for conventional age estimation methods	63
Transition Analysis scoring	64
Data analysis.....	67
Intra-observer error test.....	67
Hypothesis 1. Testing accuracy and reliability of ADBOU Age Estimation 2.1	68
Modeling age-progressive changes in the Asian samples for Hypotheses 2-4.....	69
Hypothesis 2. Evaluating sex differences in aging rates.....	78
Hypothesis 3. Population differences in aging rates	78
Hypothesis 4. Evaluating pooled-sample parameters	79
Additional Test 1. Accuracy and reliability of conventional methods	80
Additional Test 2. Retrieving age estimates by combining TA and conventional methods.	81
Chapter 5 Results.....	83
Intra-observer error test	84
Hypothesis (1) Testing accuracy and reliability of ADBOU Age Estimation 2.1.....	85
Hypothesis (2) Evaluating sex differences in aging rates	93
The log-normality assumption check.....	94
Japanese males vs. females.....	100
Thai males vs. females	101
A mixed pattern in the schedule of skeletal aging between sexes	106
Hypothesis (3) Population differences in aging rates and its impact on age estimation.	106
Univariate analysis.....	108

Quadrivariate analysis.....	117
Additional comments on side differences.....	122
Multivariate analysis: Japanese male age estimation using Thai male parameters	122
Multivariate analysis: Thai male age estimation using Japanese male parameters	123
Multivariate analysis: Japanese female age estimation using Thai female parameters	129
Multivariate analysis: Thai female age estimation using Japanese female parameters	130
Hypothesis (4) Evaluating pooled-sample parameters	134
Additional Test 1. Accuracy and reliability of conventional methods	145
Additional Test 2. Retrieving age estimates by combining TA and conventional methods	146
Chapter 6 Discussion	153
Are there any sex differences in the schedule of skeletal aging?	153
Are multiple population-specific methods an absolute necessity?	158
Is there any improvement in age estimates of the new Asian models compared to ADBOU?	162
Limitations and future directions.....	168
Chapter 7 Conclusion	173
Bibliography	175
Vita	194

List of Tables

Table 1. Correlation coefficients of age indicators to known age	31
Table 2. Sample size and composition for TA validation studies.....	59
Table 3. Sample composition of the four collections used in this study.....	63
Table 4. Abbreviated scoring format for TA (Milner and Boldsen 2013).....	66
Table 5. Kappa statistics and strength of agreement (Landis and Koch, 1977, p. 165).....	68
Table 6. Descriptive statistics for the Asian samples (n=419).....	71
Table 7. Weighted Cohen's Kappa statistics and associated p-values.....	86
Table 8. Population- and sex-specific root mean square error (RMSE, in years)	89
Table 9. Pooled-male age subgroup comparisons of RMSE (in years)	89
Table 10. Pooled-female age subgroup comparisons of RMSE (in years)	89
Table 11. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Japanese males (n=138)	95
Table 12. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Japanese females (n=45)	96
Table 13. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Thai males (n=157)	97
Table 14. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Thai females (n=79)	98
Table 15. Mean ages at transition and standard deviations (SD) of three pubic symphyseal traits	103
Table 16. Most probable scores given an exact age between 20-90 years. The four scores are in order of left and right superior apices, and left and right dorsal margins.....	119

Table 17. Japanese Male RMSE: Thai male parameters based on 4 traits	124
Table 18. Thai Male RMSE: Japanese parameters based on 11 traits	124
Table 19. Japanese female RMSE: Thai female parameters based on 8 traits	132
Table 20. Thai female RMSE: Japanese female parameters based on 4 traits	132
Table 21. Pooled male Lagrange multiplier test and pseudo- R^2 before and after stage collapsing	137
Table 22. Pooled male collapsed stages of traits for each age indicator.....	138
Table 23. Pooled female Lagrange multiplier test and pseudo- R^2 before and after stage collapsing	139
Table 24. Pooled females collapsed stages of traits for each age indicator	140
Table 25. RMSE for simulated males using TA scores (30 traits)	140
Table 26. RMSE for simulated females using TA scores (27 traits)	140
Table 27. Average widths of estimated age ranges for various models used in this study.....	145
Table 28. RMSE for simulated males using conventional methods	150
Table 29. RMSE for simulated females using conventional methods	150
Table 30. TA and conventional methods combined: Pooled males parameters (n=117), simulated males (n=200)	150

List of Figures

Figure 1. Sample distributions of Japanese and Thai males and females	71
Figure 2. Sample distributions of pooled Asian males and females	80
Figure 3. Squared error of ADBOU age estimates for pooled males	90
Figure 4. Squared error of ADBOU age estimates for pooled females	90
Figure 5. Bias of ADBOU age estimates for pooled males	91
Figure 6. A distribution of age estimates and documented chronological ages for pooled Asian males	91
Figure 7. Bias of ADBOU age estimates for pooled females	92
Figure 8. A distribution of age estimates and documented chronological ages for pooled Asian females	92
Figure 9. Stage-by-stage transition distributions of the apical morphology of the left sacroiliac joint for Japanese males and females.....	99
Figure 10. Transition distributions for the left superior apex of Japanese males and females ...	101
Figure 11. Transition distributions for the left dorsal margin of Thai males and females	102
Figure 12. The probabilities of being in each stage of the left superior apex: Japanese and Thai male comparisons.....	109
Figure 13. The probabilities of being in each stage of the right superior apex: Japanese and Thai male comparisons.....	110
Figure 14. The probabilities of being in each stage of the left dorsal margin: Japanese and Thai male comparisons.....	111
Figure 15. The probabilities of being in each stage of the right dorsal margin: Japanese and Thai male comparisons.....	112

Figure 16. The probabilities of being in each stage of the left superior apex: Japanese and Thai female comparisons	113
Figure 17. The probabilities of being in each stage of the right superior apex: Japanese and Thai female comparisons	114
Figure 18. The probabilities of being in each stage of the left dorsal margin: Japanese and Thai female comparisons	115
Figure 19. The probabilities of being in each stage of the right dorsal margin: Japanese and Thai female comparisons	116
Figure 20. Posterior distributions of a Japanese male with a combination of stages 2, 3, 3, 3...	120
Figure 21. Posterior distributions of a Thai male with a combination of stages 2, 3, 3, 3	120
Figure 22. Posterior distributions of a Japanese female with a combination of scores 3, 3, 4, 4	121
Figure 23. Posterior distributions of a Thai female with a combination of scores 3, 3, 3, 4	121
Figure 24. Two Gompertz hazard models for Japanese males. A solid line is for 1947-1995 mortality database (DB) of Japanese males. A dashed line is for documented chronological ages of the Japanese collections	124
Figure 25. A distribution of estimated age ranges, point estimates, and documented chronological ages for Japanese males	127
Figure 26. A distribution of estimated age ranges, point estimates, and documented chronological ages for Thai males	127
Figure 27. Bias of Japanese male age estimates using Thai male probit parameters	128
Figure 28. Japanese male squared-error distribution	128
Figure 29. Thai male squared-error distribution	129

Figure 30. Two Gompertz hazard models for Japanese females. A solid line is for 1947-1995 mortality database (DB) of Japanese females. A dashed line is for documented chronological ages of the Japanese collections.....	132
Figure 31. A distribution of estimated age ranges, point estimates, and documented chronological ages for Japanese females	133
Figure 32. A distribution of estimated age ranges, point estimates, and documented chronological ages for Thai females	133
Figure 33. Bias of Japanese female age estimates using Thai female probit parameters.....	134
Figure 34. The pooled-male model age estimates under a uniform prior and documented chronological ages for simulated males (a uniform prior).....	143
Figure 35. The pooled-male model age estimates and documented chronological ages for simulated males (an informed prior).....	143
Figure 36. The pooled-female model age estimates and documented chronological ages for simulated females (a uniform prior)	144
Figure 37. The pooled-female model age estimates and documented chronological ages for simulated females (an informed prior).....	144
Figure 38. The TA-conventional method combined model age estimates and documented chronological ages for simulated males	151
Figure 39. Error of the TA-conventional method combined model age estimates for simulated males	151
Figure 40. Bias of the TA-conventional method combined model age estimates for simulated males	152

Figure 41. The TA-conventional method combined model age estimates and documented chronological ages for simulated males.....	152
Figure 42. Age distributions of the current Asian sample	157
Figure 43. A score distribution of the left symphyseal relief showing Stages 3 and 4 cover a wide range of age cohorts	164

Chapter 1

Introduction

A demographic profile constructed from skeletal remains is the foundation of skeletal biology as it forms the premise for positive identification in forensic anthropology and can provide insight into the lifeways of past peoples. In order to establish a biological profile, anthropologists work under the basic principle that the biological processes of a study sample are similar to that of the reference samples used to develop osteological standards. This premise allows anthropologists to apply an age-at-death estimation standard to a new target sample. However, poor performances of aging methods applied to new target groups and the complex biology of senescence have created confusion in the field concerning the appropriate approach to establishing osteological standards. That is because potential sources of error for adult age estimation have two different origins, biological and methodological, that anthropologists cannot fully control.

The first potential source of error is biological variability in human skeletal aging. Unlike growth and development of the juvenile skeleton that are tightly under genetic control (Bogin 1999; Stinson 2012), senescence (or degeneration) of the adult skeleton is easily influenced by both intrinsic and extrinsic factors that interfere with natural biological aging. Intrinsic factors include an individual's genetic makeup, sex, and ancestry, while extrinsic factors cover a much wider range of variables, such as nutritional states, pathological conditions, and different levels of activity (Kemkes-Grottenthaler 2002). All these factors contribute to variation within and between individuals. As a result, age-related morphological changes, even in two skeletons of the same age, may manifest at different rates, and different anatomical elements of one individual may indicate different ages. Indeed, investigating the effects of extrinsic factors of skeletal aging has been difficult as one needs access to additional information for each individual, such as

occupation, repetitive habitual activities or medical history, that are not always available in reference collections. The effects of activities, and body size to skeletal aging, however, started emerging in literature in recent years (Campanacho et al. 2012; Merritt 2015; Wescott and Drew 2015).

In contrast, the effects of sex and ancestry in relation to skeletal aging have been explored by many researchers for decades. Some researchers have detected sex and/or population differences in aging rates of skeletal elements (Katz and Suchey 1989; Komar 2003; Lottering et al. 2013), whereas others have not (Konigsberg et al. 2008). Handling the “sex” variable seems to be relatively easier than ancestry: when sex differences are detected, then there will be two separate methods for each sex (Suchey and Katz 1998). If there is no sex difference, then one method will suffice (Lovejoy et al. 1985a). However, when it comes to the variable of ancestry, or population, a decision becomes more difficult, because we do not yet have a clear idea if population-specific aging rates and patterns are indeed vastly different. Some studies on age estimation have distinguished Eastern Europeans and European Americans as two distinct populations (Komar 2003), while, in other instances, they are considered as the same “whites” (Berg 2008). Conversely, there are cases when a population represents a specific geographic region (e.g. Thais) broadly generalized as “Asians” (Schmitt 2004). In order to minimize such confusion, the current project distinguishes the two terms, ancestry and population. Ancestry denotes a broad group of people that either occupy a major ecogeographical area, and/or share a relatively close ancestral origin based on a recent migration history, such as peoples of the Asian, African, European, and American continents. A population refers a smaller unit of people that reside in or from a more politically defined area (e.g. Japanese and Thais). Lastly, a sample indicates a group of skeletal remains belong to particular skeletal reference collections.

Although the importance of developing and using population-specific aging methods has been emphasized among anthropologists, what level of population-specificity one age estimation method should offer relatively underexplored. Richard Lewontin argued that variation among major geographic regions (continents) accounts for 6.3% of total human genetic variation, while variation among local populations within geographic regions (sub-continents) accounts for 8.3% of the total variation. Surprisingly, the remaining 85.4% of genetic variation has been known to exist *within local* populations (Lewontin 1972). This is not a new concept to anthropologists, as we all are aware that, in skeletal aging, there are great within-group variations as much as between group variations because of the aforementioned internal and external factors. Rather, the genetic information raises the question whether it is absolutely necessary to develop population-specific methods despite the fact that there are such greater within-population differences. This further poses a question: if there is not much variability between populations, can we pool different populations to collectively develop a more inclusive age estimation methods? More specifically, should multiple methods be developed, for example, on Japanese and Thais independently, or can we establish an aging method combining the two populations from the same continent? The key to addressing this question lies in a holistic understanding of population history as well as systematic studies of subgroups within a continent.

Particularly, a study of Asian populations, including peoples of Southeast Asia (e.g. Myanmar, Thailand, Vietnam, Laos, Cambodia, Malaysia, Singapore, Indonesia, Brunei and the Philippines) and East Asia (e.g. China, Taiwan, Korea, Japan, Mongolia), offers several merits for understanding population-specific skeletal aging. First, Asia is the largest and most populated continent potentially housing large human skeletal variations. Second, Asia has an extremely well documented population history and complex migration routes with abundant skeletal and

genetic evidence that suggests a close relationship between Asian groups. Although the initial peopling of the Asian sub-regions is still under debate, anthropologists and population geneticists have suggested possibly strong ancestral ties between Southeast (SE) Asians and East (E) Asians (Oxenham and Tayles 2006; Pechenkina et al. 2013). These studies further provide a sound justification to postulate that it might be possible to develop a single age estimation method for at least some Asian populations.

Recently, Asian reference skeletal collections have become available. While these collections have remained on the periphery of academic interest in North American and international academic scenes, collections in Japan and Thailand, that the current study is based on, have been studied for sex estimation (İşcan et al. 1994; İşcan et al. 1995; Khanpetch et al. 2012; King et al. 1998; Mahakkanukrauh et al. 2013; Nagaoka et al. 2008b; Sakaue 2004), stature estimation (Mahakkanukrauh et al. 2011), age estimation (İşcan 1987; Nagaoka and Hirata 2008; Sakaue 2006) and paleodemography (Nagaoka and Hirata 2007; Nagaoka et al. 2006; Nagaoka et al. 2008a). What needs to be problematized, however, is that a focus of these studies using Japanese and Thai collections has been possibly too regional, despite the fact that the collections combined together have great potential to be used for more holistic research, asking broader questions involving more than one population.

The second source of error is more methodologically related. Initiated by two paleodemographers' pure academic inquiry asking why a mortality profile created from written historical records is different from a mortality profile reconstructed from past cemetery populations (Bocquet-Appel and Masset 1982), anthropologists identified a series of problems associated with age estimation methods. Anthropologists realized that attaining accurate and reliable age estimates are hampered by using skeletal elements only loosely related to

chronological age, inadequate reference skeletal collections, non-negligible amount of intra- and inter-observer error, and statistical models that does not fit well to skeletal data. Twenty years later, anthropologists' rigorous efforts to improve age estimation methods produced one of the most influential publications that suggested a path forward to age estimation dilemmas (Hoppa and Vaupel 2002a). This approach argues for better quantification of qualitative skeletal traits related to improved ease of application and repeatability of methods, and that age estimation using skeletal remains needs to be based on Bayesian statistics as it provides a convenient and effective way of combining known mortality records with skeletal evidence such that the age estimates are generated with minimal bias and error.

A consensus of anthropologists for developing more accurate and reliable age estimation methods has been that age estimation needs to move beyond an experience-based art to a better science. Boldsen et al. (2002b) and Milner and Boldsen (2013) have been at the forefront of this effort. They introduced a new model for age estimation, called Transition Analysis (TA). While still based on three of the traditional skeletal age indicators, including the pubic symphysis, sacroiliac joint, and cranial sutures, the authors sub-divided each indicator into multiple components that helped increase repeatability of the method. Also, based on Bayesian statistics, TA has standalone software designed to accommodate bioarchaeological research and forensic casework. Though TA is built on a more robust statistical framework than conventional age estimation methods, the method still has a major limitation in that it does not provide flexibility to adjust the mortality information and reference samples according to researchers' needs and study populations.

Lyle Konigsberg, a biological anthropologist who is a fervent proponent of Bayesian statistics, has conducted extensive research to incorporate more robust statistical models into age

estimation, and ultimately made such flexibility possible. Particularly, Konigsberg and colleagues have shown superior results of age estimation using the uni- or multivariate ordered probit regression model over others, and offer most advanced algorithms that allow researchers to incorporate research-specific reference samples and mortality data to studies of age estimation (Konigsberg and Frankenberg 1994; Konigsberg and Herrmann 2002; Konigsberg 2015b; Konigsberg and Frankenberg 1992; Konigsberg and Frankenberg 2002, 2013; Konigsberg et al. 2016; Konigsberg et al. 2008). Nevertheless, while these sophisticated methods are freely available to enhance one's research design, these approaches have not been widely applied to a large-scale skeletal samples and studies on non-American populations.

Current study

Inspired by previous studies on Asian populations and currently available methodological approaches, the current research has two central goals. The first goal is straightforward as it is to provide information concerning how well some current age estimation methods yield accurate age estimates for Asian populations. The second goal is to help partly answer the question concerning whether a population-specific aging method should be region-specific (Thais vs. Japanese) or continental-specific (Asians). In order to achieve these goals, four skeletal collections representing part of East and Southeast Asians, namely 20th century Japanese and modern Thais, were used for the analyses.

Methodologically, this study follows an osteological scoring system developed by Boldsen et al. (2002b) and Milner and Boldsen (2013) for TA. However, an analytical approach of this project is heavily based on protocols proposed by Konigsberg et al. (2008), Konigsberg and Frankenberg (2013), Konigsberg (2015b), and Konigsberg et al. (2016) for modeling age

progressive changes of the adult skeletons. Specifically, the scores of TA and conventional methods are processed using the multivariate ordered probit regression combined with Bayesian inferences. In order to investigate population-specific skeletal aging, 33 age-related traits of TA and eight variables (left and right sides combined) of the three conventional methods (Lovejoy et al. 1985a; Meindl and Lovejoy 1985; Suchey and Katz 1998) scored from the Asian samples were analyzed using the multivariate cumulative probit regression model.

The study proposes the following four hypotheses that are expanded upon in Chapter 3.

Hypothesis (1) Based on the findings of a majority of previous studies that have shown population-specificity in skeletal aging (Katz and Suchey 1989; Lottering et al. 2013; Schmitt 2004), it is hypothesized *that the original TA will yield age estimates high in error and bias for Japanese and Thai skeletal samples because the method was originally developed on non-Asian populations*. Tests of this hypothesis will help quantify the magnitude of error and bias of TA age estimates for Asian samples, if such error exists. The quantified error and bias of TA for the Asian samples will further serve as baseline knowledge to compare age estimates derived from Asian-specific parameters that will be generated in the course of testing the next hypotheses.

Hypothesis (2) Given that female skeletons show greater variability than males due to biomechanical stress related to childbearing and child birth (Kemkes-Grottenthaler 2002; Suchey and Katz 1998), it is hypothesized that *there will be a certain degree of sex differences in age progressive changes between Japanese (or Thai) male skeletons and Japanese (or Thai) female skeletons*. This hypothesis will help determine whether sex-specific age estimation models are necessary for the two Asian populations.

Hypothesis (3) In contrast to the findings of previous studies that have shown population-specificity in skeletal aging (Katz and Suchey 1989; Lottering et al. 2013; Schmitt 2004), it is hypothesized that *an age estimation model developed on Japanese will not necessarily produce inaccurate age estimates for Thais, and vice versa*. The two rationales for this hypothesis come from the closely shared population history of Japanese and Thai populations and the nature of age indicators used in this study that are known to be relatively uninfluenced by the aforementioned intrinsic and extrinsic factors. The hypothesis will help address whether or not the Japanese and Thai samples can be pooled together to collectively develop a single age estimation model for Asians.

Hypothesis (4) Further building on the two rationales of Hypothesis (3), it is hypothesized that *age estimation models based on pooled Japanese and Thai samples will perform better than Japanese- and Thai-specific models and TA or conventional age estimation methods by generating more accurate and unbiased age estimates for the Japanese and Thai samples*. This hypothesis is based on the previous studies that have shown promising results of such Bayesian approach with improved accuracy and reliability in age estimates (DiGangi et al. 2009; Kimmerle et al. 2008; Konigsberg 2015b; Konigsberg and Frankenberg 2013; Prince et al. 2008).

Chapter organization

Chapter 2 reviews the history and development of age estimation methods starting from early interest in skeletal aging in the 16th century. The age estimation methods reviewed are restricted to three age indicators, the pubic symphysis, the sacroiliac joint, and cranial sutures, used in the current study. The chapter also presents results of validation studies, and criticisms of

conventional methods. Chapter 2 further explores the technical problems associated with conventional methods with an emphasis on statistics. The chapter also presents possible solutions suggested to improve age estimation, and explains how TA was developed and why it was chosen as a main analytical approach for the current study.

Chapter 3 reviews the population history of SE and E Asia, and explains the benefits of studying the four particular collections in Japan and Thailand. The chapter ends by elaborating upon the four hypotheses that are evaluated in this project. Chapter 4 is extensive as it describes a series of statistical methods used to test the hypotheses. The chapter explains what the multivariate ordered probit regression is and why it is chosen. It also describes how Bayes' theorem and Bayesian inferences are incorporated to obtain ages at death of the Asian samples used in the current study. The chapter also outlines the model fit test, the normality assumption check, and variable selections as well as a procedure of stage collapsing within each trait. Chapter 5 presents results of the analyses, followed by Chapter 6 that synthesizes the current study's results with previous findings in the literature. The last chapter highlights the main take-home points of the research and the implications of the current study.

Chapter 2

The Evolution of Age Estimation Methods

Age-at-death estimation of human skeletal remains has at least two important applications in skeletal biology. In bioarchaeological contexts, it is essential for establishing correct paleodemographic profiles and age-specific disease diagnoses, and provides information about our ancestral life histories. In forensic investigations, age estimation helps to identify missing persons, including victims of mass fatality incidents or human rights atrocities. In order to achieve successful outcomes from these applications, accurate and precise age estimation is crucial. Over the years, paleodemographers have first witnessed and identified serious, domino-effect consequences of inaccurate age estimates. While estimated ages are initially to build mortality profiles, this mortality information is further used to infer overall life expectancy, fertility, and population growth (Kemkes-Grottenthaler 2002). With inaccurate age estimates, certain age cohorts are unrepresented in mortality profiles, eventually leading to false interpretations of the life history of past populations.

Despite the fact that paleodemographers (and anthropologists generally) acknowledge the importance of accurate and precise aging techniques, age estimation of adult skeletal remains has been a challenging task due to difficulties in (1) finding appropriate reference skeletal collections equally representing all age cohorts, (2) detecting age-related skeletal traits that correspond to chronological ages, (3) developing an objective technique to evaluate skeletal traits, and (4) finding an appropriate statistical method to model age progressive changes in bones. Coupled with these difficulties, it has been alleged that population-level variation in skeletal aging rates and patterns combined with a lack of population-specific methods exacerbate age-at-death estimation inaccuracies.

Nevertheless, paleodemographers have sought to improve age estimation based on robust statistics using adequate reference samples and have made significant advances in age-at-death estimation of human skeletal remains. As the current project's methodological approach is designed upon the current state of age estimation, it is important to review the evolution of skeletal age estimation methods. This chapter begins by explaining steps for developing an age estimation method and reviews the history of aging method development and its contributions. The chapter also presents criticisms of these original, or conventional methods, the results of validation studies of the methods that show unsatisfactory performances, and reasons why (and where) inaccuracy and bias of the conventional methods arise. The chapter concludes by discussing recent advancements in age estimation theory and statistics.

Steps for developing an age-at-death estimation method

Developing a method for age estimation using skeletal elements can be divided into three steps. First, a researcher needs to have a reference collection with documented demographic information (i.e., known chronological age, sex and ancestry). Ideally, a “good” reference collection should represent every age group in a population without bias towards a certain age group, so that when the age distribution is plotted, it should have a uniform distribution (Konigsberg and Frankenberg 1992; Konigsberg and Frankenberg 2002). However, finding an unbiased reference collection has not been easy because, depending on its origin, each collection has a unique age distribution. For example, reference collections in the U.S. have originated from three different places. One source is the 20th century dissection room collection (e.g., the Terry Collection at the Smithsonian Institution and the Hamann-Todd collection in Cleveland). This type of collection includes individuals who died at a hospital, unclaimed bodies, and/or

criminals, often representing a lower socioeconomic group. The collections' age distribution is usually skewed to older individuals and some demographic information is not necessarily correct. A second source of reference samples is a medical examiner's office or war casualties. Such collections consist of a larger number of younger individuals than older individuals. Examples include the Los Angeles Medical Examiner's samples used for Suchey-Brooks' method or the Korean War Dead Collection of McKern and Stewart's method. A third reference sample source is a donation program that is largely biased towards older individuals. A good example is the William M. Bass Donated Collection at the University of Tennessee, Knoxville. The majority of the individuals in the collection are European American males over 60 years old. To compensate for such a void in a reference sample's age distribution, some aging methods were developed from a reference sample with multiple collections combined (Boldsen et al. 2002b; Igarashi et al. 2005; Lovejoy et al. 1985a).

The second step for developing an aging method is to identify a skeletal trait or a set of skeletal traits on a skeletal element (an "age indicator" hereafter) that undergoes metamorphosis throughout a lifespan. The lifelong changes in the age indicator are divided into sequentially, ordered categories (stages or phases). Each of these categories is assumed to generally correspond to a certain age cohort. Traditionally, different sets of skeletal traits have been categorized into arbitrarily fixed intervals in an ordinal scale. This way of defining aging patterns of the skeleton is often referred as "a phase system." Suchey and colleagues' six-phase system of the pubic symphysis and Işcan and colleagues' nine-phase system for the sternal rib end are classic examples of a phase system.

After each individual's age indicator in a reference collection is scored according to the desired scoring system, a third step is more statistically involved to assess whether the stages are

truly correlated with biological age. To do this, the stages or phases of the age indicator and documented age are analyzed to obtain correlations between the two variables. If the relationship between the two is deemed to be significant, a statistical model that best explains the variables' relationship will be fitted to the data (i.e., a relationship between trait stages and known ages). The statistical model (or an equation) for age estimation eventually becomes a standard, assuming the model validation is completed and its performance is promising for age estimation.

Commonly used statistical models for developing aging methods fall under one of the two main statistical families called frequentist statistics, which is also known as a Fisherian approach after the biggest contributor, Ronald Fisher (Kruschke 2014). Frequentist statistics is characterized by model assumptions, null hypothesis significant testing, and p-values. For age estimation, the most favored frequentist model is the uni-/multivariate linear regression (or ordinary least squares [OLS] regression) with ordered categories of an age indicator placed as an independent (or predictor) variable and known chronological ages treated as a dependent (or response) variable. After regressing ages on phases, a linear regression equation is obtained. A method is provided with descriptive statistics that includes mean age at death, standard deviations, or standard error, and some range estimates (e.g. 95% confidence intervals) for each relevant phase that serves as a “standard table.”

If we are to apply a new age estimation standard to another skeletal sample with unknown age and sex (a “target sample”), the age indicator is scored based on the standard's descriptions and the scores are plugged into the standard's equation to solve for possible age at death using simple arithmetic rules. This procedure is called the inverse calibration as one needs to “invert” the equation to estimate age at death (Aykroyd et al. 1997). Alternatively, one may simply use the provided descriptive statistics. Although this seems to be a logical procedure of

age estimation, two principle problems arise from this approach: (1) the osteological technique used to evaluate age-related changes of the bones, and (2) the statistical approach adopted to process skeletal data, which will be described in greater detail below.

Early interest in age-related osteometamorphosis

The 1900s in the field of biological anthropology is marked as an era of active development of age-at-death estimation methods. Todd (1920; 1921; 1924a, 1925a, b; 1924b), Brooks, Katz and Suchey (1990; 1986; 1998), Lovejoy and colleagues (1985a), Meindl and Lovejoy (1985), Işcan and colleagues (1984; 1985), and Stout and Paine (1986; 1992; 1992) are some of the notable names associated with age estimation methods using the pubic symphysis, the auricular surface, the ecto-cranial sutures, the sternal end of the fourth rib, and bone histology. Yet, interest in senescent patterns of the human skeleton had already begun as early as the 1500s in Europe. A Belgian anatomist, Andreas Vesalius, was the very first to document the relationship between age and cranial suture closures in 1542 (Masset 1989). Continuing Vesalius' interests, scholars in Germany (Martin 1968; Welcker 1866), Portugal (Ferraz de Macedo (1892) after Masset (1989)), and Spain (de Aranzadi 1913) investigated age-related obliteration of the cranial sutures.

Osteological metamorphosis of the pubic symphysis was noted by a number of scholars in the 1800s while they were investigating gross morphological changes in the pelvis throughout different points of a lifespan (Aeby 1858; Cleland 1889; Henle 1872; Hunter 1762; after Becker et al. (2010); Todd (1920)) They identified age-related traits of the pubic symphysis (e.g., ridges and furrows on the symphyseal face) that are still part of the modern aging methods.

Although Meckel (1820) provided the very first comprehensive description of the sacro-iliac joint (Lovejoy et al. 1985a), the joint's relationship to age progression had piqued the

interest of scholars slightly later than the cranial sutures or the pubic symphysis. Schunke (1938) found that the sacroiliac joint is covered with two different types of cartilage: the iliac side of the auricular surface is covered with the fibrous cartilage, whereas the sacral side of the auricular surface is covered with the hyaline cartilage. Schunke (1938) also described that after the third decade the fibrous cartilage of the iliac side is “roughened, furred, and frayed” (Lovejoy et al. 1985a, p.19). Notably, Sashin (1930) suggested that this process is due to an age-related increase in the proportion of fibrocartilage in the joint.

Despite growing interest in skeletal changes in relation to age, the main focus of the studies aforementioned was on soft tissues and cartilages of the skeleton. The use of age-at-death estimation method in identification did not occur until the 1920s when an “anthropologically-minded anatomist” (McKern and Stewart 1957, p.1), T. Wingate Todd, developed age estimation standards using chronological changes of the pubic symphysis and ecto-/endo-cranial suture closures in a more systematical manner.

Pubic symphysis

T. Wingate Todd developed the first standard for age-at-death estimation using the pubic symphysis. He generalized osteomorphic changes of the pubic symphysis for each age cohort and further divided life-long changes of the bone into ten discrete phases. Based on his observation that there is no apparent growth of the pubic symphysis after the age of 18. Todd (1920) examined 306 European American male skeletons after puberty (18-88 years) that were procured from dissection room cadavers from the Western Reserve University Collection. A year later, he included African American males (90) and females (approx. 22) and European American females (47) to the standard (Todd 1921).

Todd identified and defined a series of age-related traits of the pubic symphysis. Todd's traits include: (1) ridges and furrows, (2) ventral beveling, (3) dorsal and ventral margin formation and completion accompanied by (4) delimitation of upper and lower extremities, (5) rim formation, (6) bony outgrowths due to ossification of tendons and ligaments adjacent to the ventral aspect of the pubic symphysis, and (7) rim breakdown. Todd attempted to track metamorphosis in each of these traits relatively simultaneously, assuming there was a fixed set of characteristics that represent a certain age cohort. Based on this idea, he further categorized lifelong changes of the pubic symphysis into ten different age-related phases. This approach is known as a phase method that individual traits of the pubic symphysis are evaluated at the same time, so that the bone can be assigned to a single phase.

Although the contribution of the Todd to pubic symphysis aging is tremendous, the method was not entirely free from error and critiques. First, Todd excluded individuals whose symphyseal morphologies did not match his definitions, thereby failing to capture enough age-related variation in the bone. Second, the majority of the sample is over 40 years after the major symphyseal changes are completed. Third, Todd's method became more difficult to apply when the pubic symphysis did not conform to the suite of traits described in the definitions. Fourth, the method provides age ranges that are too narrow (e.g. 18-19 years for Phase 1). Fifth, the method did not identify any ancestral or sex differences in the aging process. Lastly, the ages at death of the individuals in the Western Reserve collection are not correctly reported (Katz and Suchey 1986).

Thus, Todd's method was criticized for its tendency to overestimate the age of young individuals and its inaccuracy for older individuals (Hanihara and Suzuki 1978; McKern and Stewart 1957). Brooks (1955) tested Todd's method on a series of 194 male and 177 female

archaeological skeletons from California, and a sample of 103 males and 82 females from the Western Reserve University Collection (the same collection that Todd used for developing his methods). For age estimates of individuals over 20 years, Todd's phases consistently yielded a higher than actual age. Brooks further claimed the pubic symphyseal morphologies that Todd identified as "extreme deviants" are not necessarily anomalous, but rather alternative patterns of the symphysis.

Realizing that Todd's method does not adequately accommodate individual variation in pubic symphysis morphology and has a tendency to overestimate the age of young individuals, McKern and Stewart (1957) developed an alternative approach for pubic symphysis age estimation with two specific goals in mind: (1) to provide a better method that accommodates variation in osteological metamorphosis; and (2) to provide a better age-at-death estimation standard that minimizes inter-observer errors. The standard of McKern and Stewart is comprehensive, encompassing age-related changes of the pelvis, cranial sutures, clavicles, vertebrae and the sternum.

McKern and Stewart developed their method based on American soldiers killed in the Korean War (n=450) repatriated from North Korea in "Operation Glory" during 1954. As most of the U.S. soldiers were young, the ages of death of 18-20 were most frequent. The entire range of the individuals in the collection is 17-50 years. The majority of the sample is of European descent (90.4%), and 9.3% is of African descent with only one Asian. The authors' most notable work is that, to incorporate variation, they adopted a component scoring system. Before McKern and Stewart, the pubic symphysis was evaluated based on a phase system that requires observers to identify a specific suit of skeletal traits that supposedly represent a particular age. As a result, observers are forced to choose a universally fixed, single phase without flexibility to incorporate

variation when the pubic symphysis shows traits that deviate from a particular age group. A component scoring system, on the other hand, requires observers to evaluate and score each of the traits independently, further allowing researchers to incorporate variations in symphyseal traits.

Starting from the pubic symphyseal traits that Todd identified, McKern and Stewart narrowed down the most age-informative traits to three: (1) the dorsal plateau, (2) the ventral rampart, and (3) the symphyseal rim. Each of these three traits is further evaluated based on multiple developmental and senescent stages ranging from zero (absence of the trait) to five, and receives a score independently. Although McKern and Stewart's method indeed improved Todd's method, it still had limited applications because the method was developed to analyze specific a specific demographic: young white American male soldiers.

Brooks and Suchey (1990), Katz and Suchey (1986), and Suchey and Katz (1998) also revised Todd's method based on a modern, large-scale sample (n=1,225) representing various ancestral groups, including European Americans, African Americans, Mexican and Asians. The pubic symphyses were procured from autopsy cases at the LA Medical Examiner's Office during 1977-1979 with an age range of 13-99 years old. After an extensive, five-year inter-observer error study, Brooks and Suchey (1990) concluded that a number of phases in the Todd system should be collapsed because the observers could not differentiate them. As a result, the authors presented a six-phase method by collapsing Todd's Phases 1, 2, and 3 into one phase as well as combining Phases 4 and 5 and Phases 7 and 8.

Another notable refinement of the Suchey-Brooks' method is that it captures the wide variability of the female pubic symphysis (n=273) by oversampling individuals of the same age (e.g., 20 females at the age of 28 years). Their study has shown that individuals develop and age

at different rates producing variation in symphyseal morphology. The authors also noted that Phase 1 and 2 provide the best age at death estimation with little overlapping of estimated age ranges. However, the Suchey-Brooks' method becomes problematic from Phase 3 and above as there is significant overlap of age estimates (e.g. Phases 3 and 4 have age ranges of 21-46 years and 23-57 years, respectively for males). Thus, individuals in Phase 6 are lumped into a broad age range of 34-87 years with a mean age of 60 or 61 years.

Realizing Suchey-Brooks' method cannot specify individuals' ages at death above Phase 6, Berg (2008) proposed to add Phase 7 at least for females in their mid 70s that typically show lightweight bone with extensive erosion of the symphyseal rim due to osteoporosis and osteopenia. Likewise, Hartnett (2010) presents a revised Suchey-Brooks' method with Phase 7 added for individuals over 70 years, but the age ranges of the proposed Phase 7 significantly overlap with the Phase 6 of the Suchey-Brooks' method.

Sacroiliac joint

Following the previous study of Sashin (1930) on the sacroiliac joint, Lovejoy et al. (1985) characterized distinct patterns of the auricular surface of the ilium and its posterior aspect into eight phases with a five- to ten-year interval. Lovejoy et al. (1985) used three different collections including approximately 500 individuals from the Todd collection, over 250 individuals from the Libben archaeological collection, and 14 forensic cases.

According to their pioneering descriptions on the sequential changes of the auricular surface, young auricular surfaces have a fine granular appearance with transversely organized billowing and striae. As one becomes older, there is a decrease in granularity and increase in densification of the auricular surface. There is also increased porosity with age. In the young, the

apex is sharp and well-defined, while it becomes more irregular and blunt with age. The posterior aspect of the auricular surface, the retroauricular area, is smooth in young individuals, but as one ages bony exostoses develop on this area where lumbosacral and sacroiliac ligaments are located, giving the area a more irregular appearance.

Lovejoy et al. (1985a) found no sexual dimorphism in the aging process of the sacroiliac joint, and there was a similar level of inter-observer error in the method as found for the pubic symphysis. Although Lovejoy et al. (1985a) championed the utility of the auricular surface for age estimation due to its better preservation rate than other skeletal elements, they noted that the method yields more accurate age estimates for young individuals than the old. Thus, their method can only be accurately applied when one can seriate multiple auricular surfaces simultaneously. Therefore, the method was more suitable for analysis of population-level, archaeological samples than individual forensic cases. Nevertheless, the method of Lovejoy et al. (1985a) suffered from several issues: inability to incorporate inter-individual errors, overly too narrow age ranges, and open ended beyond 60 years (Buckberry and Chamberlain 2002).

Decades later, following McKern and Stewart's approach, Buckberry and Chamberlain (2002) revised Lovejoy et al.'s (1985a) method based on a component-scoring system. Each of the traits were scored independently, summed, and compared to the provided standard table to obtain final age estimates. The authors used 180 individuals from a documented 18th-19th century English cemetery (the Spitalfields Collection). Buckberry and Chamberlain (2002) eliminated traits of the original method that were deemed to be not as age-informative as Lovejoy et al. (1985a) initially claimed. Of the original six traits, the retroauricular activity was eliminated from the revised method because it did not perform well for age estimation with their sample.

The work of Buckberry and Chamberlain (2002) is notable in a sense that they elaborated upon Lovejoy et al.'s method by combining a component-scoring system on an ordinal scale to make the method easier to apply with low inter- and intra-observer error. The authors found no significant difference between sexes and sides, allowing the same method and age estimates to be applied to both sexes. Yet, the authors do recognize the limitation of their study: a small sample size ($n = 180$) with younger individuals underrepresented. The method provides wider age ranges than Lovejoy et al.'s method, but, in fact, are too broad to be utilized for identification purposes.

Osborne et al. (2004) presented another notable revision of Lovejoy et al.'s (1985a) method. The authors tested the method using a target sample different from Lovejoy's reference sample, and the results showed that the method performed with only 33% accuracy. Using the Terry and Bass collections ($n=266$, 16-89 years), the authors concluded that their new six-phase system performed better than Lovejoy et al.'s (1985a) original eight-phase system. Osborne et al. did not find any sex or ancestry specificity or secular change in aging process, though admitted that age only explained a small amount of variation seen in the auricular surface (the adjusted r^2 value= 0.34). Osborne et al. concluded that the auricular surface is more age-informative than the pubic symphysis since the auricular surface undergoes prolonged metamorphosis throughout one's lifespan, whereas changes in the pubic symphysis are no longer obvious once the ventral rampart is completely formed (during the third or fourth decade of life).

A year later, Igarashi et al. (2005) presented an age estimation method using the auricular surface of Japanese skeletal samples. The authors used 700 modern documented Japanese skeletal remains (438 males and 262 females) with an age range of 16- 89 years. To reduce intra-/inter-observer error, the authors developed a method based on simple presence or absence of 13

different variables. The variables they included are related to surface relief (wide groove, striation, roughness, flatness), texture smoothness, fine granularity, coarse granularity, sparse porosity, dense porosity), and hypertrophied bony structure (dull rim, lipping, tuberosity, bony bridge). Perhaps the most notable finding of their work is that there is a certain combination of auricular surface traits that works better for males and females separately. For example, including fine granularity, dense porosity, and tuberosity in their multiple regression model yielded the best age estimates for males. For females, adding fine granularity and dense porosity provided the most accurate age estimates.

Continuing revisions of the component-based method of the auricular surface, Falys et al. (2006) presented a revised method of the Buckberry and Chamberlain (2002) using a documented English skeletal collection from the 17th-19th century. Perhaps, their results are most disappointing of all other studies on the auricular surface. The authors reduced the original seven phases to three phases with even broader age estimates that span almost the entire adulthood. In their concluding remarks, the authors state that there is a little hope for age estimation using the auricular surface, and the method is not suitable for forensic purposes.

Cranial sutures

The literature on age-related synostosis of cranial sutures far exceeds any others, and even goes back to the time of Hippocrates (Masset 1989). Unlike the early studies of the pubic symphysis and the auricular surface, various populations with different ancestral backgrounds were explored, and sex and ancestral differences in suture closures were noted from the skulls of Americans, Hungarians, Mexicans, Romanians, Czechs, and South Africans (Masset 1989; Singer 1953). Yet, many of the early authors concluded that suture closure is extremely variable

resulting in unreliable age estimates, and that therefore there is a little to no value of using cranial sutures for identification (Bocquet-Appel and Masset 1982).

Todd and Lyon (1924a, 1925a) revisited age-related synostosis of both ecto- and endocranial sutures using the Western Reserve University Collection (n=307 European males, and 120 African males) in an attempt to resolve the issue of unreliability of suture closure for age estimation. However, their results did not improve upon the earlier deficiencies; there is too much variability in progress of cranial suture closure to be age-informative. Therefore, should suture closures be used as an age indicator, then final age estimates must be drawn in junction with other age indicators.

More than five decades later, Meindl and Lovejoy (1985) presented an aging method based on ectocranial sutures using 236 crania from the Hamann-Todd Collection. The method involved 10 points of sutures (each 1 cm long) from the vault and the lateral-anterior skull. The vault region has seven points to score: midlambdoid, lambda, obelion, anterior sagittal, bregma, midcoronal, pterion. The lateral-anterior region has five points to score: sphenofrontal, inferior sphenotemporal and superior sphenotemporal, pterion and midcoronal. The scoring scale ranges from 0=open and 3=complete obliteration.

Meindl and Lovejoy found the strongest correlations with age in four cranial sites, including: pterion, sphenofrontal, midlambdoid and lambdoid. Thus, the lateral-anterior system was found to yield better estimates than the vault, especially for older ages. Their ectocranial suture method gives better estimates than a previous endocranial method (Acsádi and Nemeskéri 1970). Their study found no significant differences by ancestry or sex. Meindl and Lovejoy (1985) argued that this method provided more accurate age estimates than those of the pubic symphysis. The authors also suggested that fully open crania should be aged based on postcranial

indicators whereas fully closed crania should be aged based on pelvic and radiographic indicators and dental wear.

In following years, a number of researchers have attempted to improve age estimation using cranial sutures (Dorandeu et al. 2008; Masset 1989; Perizonius 1984). Most recently, Alesbury et al. (2013) sought to utilize frontonasal sutures for age estimation, but the results were rather disappointing and age only accounted for 13% of suture variation. Regardless of the disappointing age estimates using cranial sutures, it is still one of the most commonly used in the field as often it is the only age indicator available in forensic cases or archaeological human analyses, and the observer is left with no alternative.

Validation studies

As the most-used conventional age estimation methods were developed exclusively on North American samples, subsequent validation studies of the aforementioned age estimation methods have been conducted since the 1980s. Although there still is a large void in the literature concerning studies on Asian groups, various populations were used for these validation studies, including modern Americans (Martrille et al. 2007; Meindl et al. 1985a; Murray and Murray 1991), Canadians (Bedford et al. 1993; Saunders et al. 1992), French (Baccino et al. 1999), British (Aiello and Molleson 1993; Falys et al. 2006), Spanish (Rissech et al. 2012), Serbians (Djurić et al. 2007), Indians (Sinha and Gupta 1995), Thais (Schmitt 2004) and Australians (Lottering et al. 2013). Target sample sizes used in the validation studies varied from as little as 19 individuals up to 296 individuals, who were procured from archaeological and/or forensic settings. These validation studies showed that the average inaccuracy ranges from 3-15 years, and the average accurate classification rates range from 27-90%.

Based on these results, the majority of the authors of those validation studies were skeptical and disappointed by the performance of the tested methods (Falys et al. 2006; Schmitt 2004), and a few supported the utility of conventional aging methods if multiple methods combining micro- and macroscopic techniques are going to be used (Aiello and Molleson 1993; Dudar et al. 1993). These validation studies further identified five general trends in inaccuracy and bias of the conventional methods. First, age estimation is most accurate for young individuals, and there was an increase in inaccuracy and bias of the method with older age groups. Second, systematic over-estimation of the young and under-estimation of the old persist. Third, conventional methods consistently fail to capture individuals over 50 or 60 years, thus mortality profiles falsely demonstrate that no one lived beyond 50 or 60 years of age. Fourth, the conventional methods fail to quantify bias or error rates associated with their estimated ages. Fifth, it is unclear whether the methods developed on the American population are equally applicable for a target sample from a different population. In addition, age indicators that were traditionally believed to be most reliable (e.g. the pubic symphysis) were not free from these problems, and showed the same increased inaccuracy and bias as age increases.

To explain such unsatisfactory performance of aging methods, ancestral or population differences in skeletal aging process has always been considered as the basic, “by-default” factor that increases inaccuracy of age estimates (Katz and Suchey 1989; Komar 2003; Lottering et al. 2013). Nonetheless, there are many other biological and non-biological factors that may cloud clear assessment of natural biological aging process of age indicators, thus contributing to poor performance of aging methods. Those obstacles include different nutritional states, repetitive activities, pathology, secular change, sampling bias, and inter-/intra-observer errors, to name a few (Kemkes-Grottenthaler 2002; Usher 2002). Although these do, in part, contribute to

increased inaccuracy of age estimates, these reasons alone do not fully explain the trends. A complete explanation on what causes these issues comes from the statistical approach underpinning the conventional aging methods. In the following sections, problems with the osteological and statistical technique that the conventional methods are based on to evaluate age-related changes on bones, and the importance of using appropriate statistical approaches to analyze skeletal data will be discussed.

Problematic osteological techniques of conventional age estimation methods

Conventional age-at-death methods are problematic in their application to unknown skeletal samples because of three important technical and statistical reasons. The first technical problem of conventional methods is that they utilize a single bone element, or “age indicator,” as their single source of information for age expression. They are, thus, unable to capture variation in aging rates (or inconsistencies) expressed in different parts of a single skeleton (within-individual variation) (Baccino et al. 1999; Boldsen et al. 2002b; Buckberry and Chamberlain 2002; McKern and Stewart 1957). Consequently, unless an observer uses various age-at-death estimation methods, and collectively summarizes different age-at-death estimates from multiple age indicators, the observer is likely to make a cumulative age-at-death estimate that does not take into account the different rates of aging within the skeleton of the same individual.

Second, most conventional age-at-death estimation methods do not incorporate variation among individuals within a population (between-individual variation) (Boldsen et al. 2002b; Buckberry and Chamberlain 2002; McKern 1957) unless the method is based on a component-scoring system, such as the methods of McKern and Stewart (1957) and Buckberry and Chamberlain (2002). It has been reported that both genetic predisposition and external factors

(e.g. nutrition, biomechanical forces, trauma, and pathologies) influence how skeletal traits are manifested, and therefore no two skeletons of the same age demonstrate identical age-related skeletal characteristics (Kemkes-Grottenthaler 2002). Nevertheless, because conventional methods require observers to evaluate a suite of age-related skeletal traits on a single age indicator, including texture, topography and porosity, an individual skeleton is forced into a universally fixed age category (or phase). As a result, conventional methods do not capture the different aging rates and patterns between individuals. Moreover, because of this, conventional methods are notoriously difficult to apply to fragmentary age indicators.

The third technical problem of conventional age-at-death estimation methods comes from the subjective nature of its application (Hoppa and Vaupel 2002b; Jackes 2000; Kemkes-Grottenthaler 2002). Most descriptions given in conventional methods use unquantifiable, vague terminologies, such as ‘slight,’ ‘minimal’ or ‘being in good condition’ to describe skeletal traits that can be interpreted in vastly different ways from one observer to another (Kemkes-Grottenthaler 2002). This can result in significant inter-observer error even when the observers are using the same method. It can also manifest as intra-observer variation when repeated assessments are made on the same skeleton at different times.

To resolve these problems with conventional methods, alternative approaches were suggested (Bedford et al. 1993; Jackes 1985; Lovejoy et al. 1985b; Meindl et al. 1985b; Meindl et al. 1990). Specifically, instead of using single age indicators, Lovejoy et al. (1985b) introduced a multifactorial method that combines multiple age indicators (the pubic symphysis, auricular surface, radiographs of proximal femur, dental wear, and suture closures) based on a principal components weighting. To reduce intra- and inter-observer error and improve repeatability of the method, Igarashi et al. (2005) simplified skeletal traits into a binary (presence

or absence) scoring system. Nevertheless, these methods have not been extensively applied because, in part, obtaining final age estimates is not as simple as those of conventional methods that come with standard tables with age estimates for each phase. Most importantly, these alternative methods did not fully offer a resolution to problems rooted in the statistical model from which the conventional methods were developed.

Statistical problems with conventional aging methods

Influence of reference collections on the age-at-death distribution

Error associated with the statistical approach of conventional methods was first addressed by Bocquet-Appel and Masset in their 1982 publication when they bid “Farewell To Paleodemography.” What had inspired them to initiate systematic investigation on aging methods was the fact that there were non-negligible discrepancies between documented death rate of historical demography data and estimated ages at death of cemetery skeletal samples. Specifically, while age-at-death distributions of the latter suggested a short lifespan of the past population, age distributions of the former showed a longer lifespan. Soon, their investigation revealed that the discrepancies originate from an erroneous statistical approach underpinning most of the conventional aging methods, not due to anthropologists’ malpractice or biological nature of human skeletal variation.

The most problematic issue that Bocquet-Appel and Masset (1982) detected was the fact that, regardless of the aging methods used, estimated ages of a target sample closely resembles the age distribution of the reference sample used to develop the method. The authors define this phenomenon as “the procedure which leads one to superimpose on the mortality structure of cemetery populations the structure of other populations entirely alien to them (p. 321).” The

authors find the reason for this phenomenon from the use of the OLS regression, a simple regression that processes known ages as a dependent variable and skeletal traits as an independent variable. This way of data processing (i.e. the inverse calibration) is not only biologically incorrect—we do not age because skeletal traits change, but rather the skeleton changes because we age—but also ultimately produces age estimates sensitive to the age-at-death distribution of the reference collection. The authors demonstrate the substantial influence of the reference samples with a side-by-side comparison of an estimated age distribution of the 11-13th century Sudanese Nubians using the McKern and Stewart's method of the pubic symphysis (1957, n=373). The two different skeletal samples showed very similar age structures between the ancient Nubians and the McKern and Stewart's reference sample because the OLS regression imposes the age structure of the reference sample onto the target samples.

The method of McKern and Stewart was developed on a skeletal collection of American soldiers who died during the Korean War. Naturally, the reference sample is heavily biased towards males in their late teens and 20s, consequently better capturing skeletal characteristics of young adults, more than middle- or older-aged adults. Therefore, when the McKern and Stewart method is applied to an unknown target skeleton of, for example, a fifty-year old, it will likely underestimate age, even if the skeleton shows characteristics of a typical middle-aged adult. Moreover, if the method is applied to a large number of skeletons, the estimated sample age distribution will “mimic” the Korean War dead collection, failing to reflect the actual age distribution of the target samples at interest. This phenomenon was later named “the phenomenon of age structure mimicry” by Mensforth (1990, p.91) and is now simply referred to as “age mimicry.”

Another crucial point that Bocquet-Appel and Masset (1982) emphasize is that the phenomenon of age mimicry is exacerbated when there is low correlation between an age indicator and known age. The authors argue that, in order for an age indicator to be reliable, a minimum correlation coefficient between age indicators and known ages should be 0.9 or greater. Konigsberg and Frankenberg (1992) explain why this is the case by presenting two extreme, unlikely values of correlation coefficients between an age indicator and known ages: A zero correlation and a perfect correlation. When an age indicator has no relationship with known ages ($r^2 = 0$), the age indicator will be completely independent from chronological age. This independence further boils down a model equation in a way to reproduce the reference sample's age distribution (Konigsberg and Frankenberg 1992, p.238). As a result, a target sample's estimated age distribution will be exactly as same as that of the reference distribution. On the other hand, if an age indicator is perfectly correlated with age ($r^2 = 1$), then the age estimates will be completely free from age mimicry (Konigsberg and Frankenberg 1992). Although both the extreme cases are unlikely to exist, it is as rare to find an age indicator with a correlation higher than 0.9 (Table 1).

Understanding the etiology of age mimicry further unveils the reason why conventional methods systematically overestimate the age of young adults and underestimate that of old adults, and why the elderly individuals are often invisible in estimated age-at-death distributions (Aykroyd et al. 1999; Aykroyd et al. 1997; Masset 1989). First, the OLS model regresses age on skeletal traits to create a linear line (the model equation) that has a minimum error of distance from each of the data points (age indicator stages) that will always produce older estimates for young individuals and younger estimates for old individuals (Aykroyd et al. 1997; Masset 1989).

Table 1. Correlation coefficients of age indicators to known age

Age Indicator	Method	Female	Male	Both Sexes
Pubic Symphysis	McKern and Stewart ¹		0.72	
	McKern and Stewart ²	0.68	0.37	0.36
	Todd ¹		0.85	
	Todd ²	0.68	0.37	0.36
Sacroiliac Joint Surface	Meindl et al. (1985b)			0.72
	Buckberry and Chamberlain (2002)			0.89-0.94
	Osborne et al. (2004)			0.59
Ectocranial Sutures	Meindl and Lovejoy ³			Lateral- 0.57, Vault-0.50
	Meindl and Lovejoy ⁴	0.34	0.59	0.56
Multifactorial	Lovejoy et al. ⁴	0.79	0.90	0.85
	Boldsen et al. (2002b)			0.88

References: ¹ Katz and Suchey (1986); ² Meindl et al. (1985b); ³ Meindl and Lovejoy (1985);
⁴ KemkesGrottenthaler (1996) (The table reproduced after Kemkes-Grottenthaler (2002))

Second, older individuals are missing from estimated age distributions because a regression model established from a reference sample “drifts” the estimated age distribution of a target sample towards the mean of the reference sample distribution (Konigsberg and Frankenberg 1992). This trend, in part due to using the inverse regression model, is named as “the attraction to the middle” by Masset (1989). This is not entirely a new finding as statisticians have identified it as “regression (or reversion) to the mean” since Galton (1973) (Konigsberg and Frankenberg 1992, p.249). Here, having a high correlation between an age indicator and known age plays a major role once again. When an age indicator has a lower correlation to known age that is less than one, not only will the estimated age distribution of the target sample resemble the reference sample distribution, but also the estimated mean age of the target sample will be biased toward the mean age of the reference sample (Konigsberg and Frankenberg 1992).

Based on their analysis, Bocquet-Appel and Masset (1982) argue that historical mortality records may be more reliable than age-at-death distributions derived from skeletal collections. And yet, the authors warn researchers that historical records are also not free from bias and

manipulation, further concluding any research on paleodemography is not recommended. This reverberating publication (Bocquet-Appel and Masset 1982) resulted in innumerable debates and response papers (Bocquet-Appel and Masset 1985; Bocquet-Appel 1986; Buikstra and Konigsberg 1985; Greene et al. 1986; Konigsberg and Frankenberg 1994; Lanphear 1989; Masset 1989; Piontek and Weber 1990; Van Gerven and Armelagos 1983).

Traditional statistical approaches and their assumptions

Aside from the issue of accuracy and reliability of age estimates, many authors have questioned the rationale for choosing a statistical model to develop an aging method. Particularly, Boldsen et al. (2002b); Konigsberg et al. (2008); Lucy et al. (1996); and Lucy and Pollard (1995) have addressed basic assumptions of linear regression models that skeletal data do not necessarily satisfy. Although violating some of the assumptions in the following paragraphs does not necessarily lead to inaccurate age estimates, there is no justifiable rationale for taking such statistical approaches (Lucy et al. 1996).

The first assumption is that, *variables are continuous and normally distributed*. However, this assumption is violated upfront given that most traditional aging methods employ a handful of ordered categorical stages or phases that are far from a continuous scale. When a variable is normally distributed, it also means that its error distribution is normal (Konigsberg et al. 2008). However, there is no supporting rationale that age-at-death distributions of humans are normally distributed (Konigsberg et al. 2008). Konigsberg further adds that it is entirely not convincing to believe that an age-at-death distribution within each phase is symmetrical or behaves as a Gaussian distribution. A violation of the normality assumption *may* influence model prediction and results of the regression due to non-optimal regression coefficients, but *will* result in non-

normal distribution of error around its predictive value that further affects reliability of interval estimates, hypothesis testing and regression parameters (Quinn and Keough 2002).

Second, *all errors are only in the direction of the response variable, y, which is, conventionally, known age* (Lucy et al. 1996). Aykroyd et al. (1997) explains that the error actually comes from age indicator scores (predictor variable) as well, because each score can be assigned to a possible age range, not an exact age, due to biological variation and uncertainty. Lucy and Pollard (1995) suggest that, instead, it is more plausible to choose the known age as an independent variable, x, and skeletal traits as a dependent variable, y, assuming that “errors associated with rounding the age to whole years” is minimal (p. 224).

The third assumption is that, *all variables provide information on age independent from each other*. If this assumption is true, a trait on an age indicator cannot cause changes in another trait for the same age indicator (e.g., a change in the ventral rampart of the pubic symphysis cannot change the dorsal plateau of the same bone). As a matter of fact, multiple age indicators can be correlated with one another (Boldsen et al. 2002b). The issue of multicollinearity can be treated either by assuming conditional independence (Boldsen et al. 2002b; Lucy et al. 1996) or by establishing a correlation matrix so that the age indicators that are highly correlated to each other can be manually weeded out (DiGangi et al. 2009; Konigsberg et al. 2008).

The fourth assumption, in the case of a multifactorial or summation approach, is that, *all age indicators have an equal contribution to age estimation*. This means that two different traits of one age indicator provide the same amount of information on age at death. Yet, this assumption is not necessarily true given that some skeletal traits have higher correlation with chronological age than other traits (Lucy et al. 1996). Regardless, many traditional aging

methods are based on this untenable assumption since the methods require a simultaneous evaluation of multiple age-related traits and give a single phase.

The fifth assumption is that, discrete age intervals (or phases) indicate that *every skeleton has the same amount of error or precision* (Boldsen et al. 2002b, p.75). That is, a researcher is likely to make the same mistakes on estimating age for a 20 year old as for a 50 year old. This assumption is unsupportable as everyone has an individually unique skeleton, and variability in trait expressions between individuals increase as one becomes old. This gives a different amount of error and bias in age estimation from one skeleton to another. Some of the factors that may influence the chances of researchers misclassifying age include; preservation condition of the bone, or premature aging due to diseases, strenuous physical activity, nutritional condition, etc. For this reason, it is important to present uncertainty in age estimates that is specific to a skeleton being analyzed.

Realizing the drawbacks of using the linear regression and equally spaced phases, Katz and Suchey (1986) present a modified Todd's method of male pubic symphyses, and advocate using the percentile method. Usually, the end products of the percentile method consist of a mean, a standard deviation and an age range for each age-indicator phase (or stage) that categorize a certain age cohort (c.f. Katz and Suchey (1986)'s Table 8). Though it is a common practice by current anthropologists, Konigsberg et al. (2008) argue that it is "perilous" to directly apply the summary statistics (e.g. a mean age and SD) to unknown target samples, because the percentile approach performs no better than the OLS regression in eliminating 'the age estimation problems.' First, sampling variance can be substantial and the standard error (SE) within a phase can be too large if each of the phases contains a few individuals that do not adequately represent the phase. However, such large variance and SE has been defended in the

literature as incorporating much variation and presented as preferable (Konigsberg et al. 2008). Second, broad age ranges of summary statistics of a percentile method is not always informative for anthropologists, because its estimates span much of the adulthood, further making anthropologists difficult to narrow down possible matches. Third, traditional percentile methods contain an implicit prior distribution for age (a.k.a. hidden Bayesian). This implicit prior is the actual age distribution of the reference sample. Therefore, unless a researcher can justify that the age structure of a target sample is similar to that of the reference sample, issues of age mimicry will result.

In a broader statistical context, age mimicry and its associated problems are discussed as “the calibration problem” (Konigsberg and Frankenberg 1992; Konigsberg et al. 1997; Konigsberg et al. 1998; Masset 1989). By definition, calibration is “the process of assigning values to... the property of an artifact relative to reference standards or measuring processes (Cameron 1982, p.346, after Konigsberg et al. 1997, p.67)” According to Konigsberg et al. (1997, p.67), anthropologists use a modern reference sample (also, a calibration sample) with known ages to assign ages-at-death to unknown target skeletons based on their “properties,” which is one or more discrete, ordinal categorical or continuous indicators that are dependent on age. In the inverse regression, known age is regressed on the indicator in a calibration sample. That is, known age is treated as the dependent variable, while the indicator is placed as the independent variable. As briefly mentioned in the steps for developing an age estimation method, this statistical approach is most commonly used. Using the inverse calibration, however, is not entirely problematic as long as one can justify that age distributions of both target and reference samples are similar before applying conventional methods. If this cannot be justified, age estimates will be subject to the previously discussed issues.

The classical calibration, on the other hand, regresses the indicator on known age followed by solving for age. Using the classical calibration eliminates age mimicry, and therefore, it is favored over the inverse calibration (Aykroyd et al. 1997, Aykroyd 1999). Although the classical calibration yields unbiased age estimates, it is known to be inefficient, and the estimates are not exactly what we want, because the model predicts probabilities of observing skeletal traits given age, not vice versa. In response to this problem, Konigsberg et al. (1997) and Konigsberg and Frankenberg (1992) recommend using the maximum likelihood method for age estimation that is concordant to the spirit of classical calibration, thus providing more efficient estimates.

Inspired by fisheries literature (Bartoo and Parker 1983; Clark 1981; Fournier and Breen 1983; Kimura 1977; Kimura and Chikuni 1987), Konigsberg and Frankenberg (1992) introduce an “iterated age length key” method with the Bayes’ rule as a “more elegant way” to treat the problem of age structure estimation based on morphology. The field of fisheries has had a longer history of dealing with the phenomenon of age mimicry (Kimura 1977). The iterated age length key, a method used in fishery literature to estimate an age of fish based on its body lengths is “an expectation-maximization (EM) algorithm (Dempster et al. 1977) that leads to maximum likelihood estimates for the age-at-death structure for a target sample (Konigsberg and Frankenberg 2002, p.498).” In their 1992 paper, Konigsberg and Frankenberg prove that the maximum likelihood method does not generally suffer from biased age estimation, and therefore is recommended over the OLS regression. A decade later Konigsberg and Frankenberg (2002, p.306) explicitly state that using Bayes’ theorem is “the only logical way to proceed in estimating age in paleodemography.”

A possible solution: Bayes' Rule/Theorem

As proposed by Konigsberg and Frankenberg (2002), Bayes' theorem indeed provides several benefits to anthropologists who wish to develop an age estimation method or to apply an existing method to a new target population. The most advantageous aspect of adopting Bayes' theorem is that it allows us to effectively combine both biological evidence (skeletal elements) and documented information of death (the age-at-death distribution of the population at interest) to achieve unbiased and more reliable age estimates. In the broadest sense, Bayesian statistics allows researchers to compute the unobservable (e.g. age at death) based on the observable (e.g. skeletal evidence).

The core function of Bayes' theorem is to update prior knowledge on probabilities of a certain event occurring based on a new set of evidence (e.g. data from study samples). The updated prior knowledge corrected by new evidence becomes the probability that an observer ultimately wishes to know (a posterior probability). In the context of age estimation of an unknown skeleton, we begin with a random chance of dying for each age of a target population that is often referred to as a known age-at-death distribution of the target population. This distribution can be obtained from various sources, including national census data on deaths, homicide databases, historic parish records, etc., and eventually becomes our prior knowledge. This prior knowledge, or the prior probabilities in Bayesian terms, provides the probability of the unknown skeleton's age before any osteological observations are made. A prior can be a specific distribution (an informed prior) or a vague, non-informative prior (a uniform, uninformed, or flat prior), which assumes everyone has an equal chance of dying at each age. Once the most relevant prior to the case at hand is determined, new sets of information are gathered that will be age-informative traits from a reference skeletal sample that further become likelihood functions, to

update our initial knowledge on the age at death. In the end, the updated prior becomes the posterior probability (final age estimates). The posterior probability can be a point estimate if the prior probability is a point value. However, for age estimation, it is recommended to use a prior “distribution” to obtain a posterior probability “distribution” in order to presents our uncertainty in age estimates (Boldsen et al. 2002b). Alternatively, these steps of Bayesian inference lead to a simpler form of the Bayesian theorem in the below Equation 1 that the posterior probability is proportional (not equal) to a product of the likelihood and the prior probability:

$$\text{Posterior Probability} \propto \text{Likelihood} \times \text{Prior Probability} \quad (1)$$

The same year of 2002, when Konigsberg and Frankenberg (2002) proposed the Bayes’ theorem as a solution to age estimation dilemmas, marks the full paradigmatic shift to Bayesian statistics in the context of age estimation. The most notable event is the publication of *Paleodemography: Age Distributions From Skeletal Samples* (Hoppa and Vaupel 2002a). This seminal volume is not only a product of a three-day workshop held in Rostock, Germany, in 1999, but also stands as a scholarly consensus of three decades of efforts to develop better aging methods based on sound statistics and effective skeletal evaluation techniques. As a solution to improve age estimation, Hoppa and Vaupel (2002b) put forth “the Rostock Manifesto” with the following four key elements (p.2), which is essentially an expansion of the simple Equation 1 above:

- 1) Osteologists must develop more reliable and more vigorously validated age indicator stages or categories that relate skeletal morphology to known chronological age.
- 2) Using these osteological data, developed models and methods should estimate *the probability of observing a suite of skeletal characteristics (c) given known age (a), which is noted as $P(c|a)$* . This probability, or likelihood functions, comes from the reference sample.
- 3) However, what osteologists want to estimate is *the probability that the skeleton is from a person who died at a certain age (a) given the evidence, the skeletal characteristics (c)*.

This probability is noted as $P(a|c)$, and is not equal to $P(c|a)$ in (2). Therefore, in order to invert $P(c|a)$ to $P(a|c)$, Bayes's theorem, Equation 2, must be applied.

$$P(a|c) = \frac{P(c|a)f(a)}{\int_0^w P(c|x)f(x)dx} \quad (2)$$

- 4) Using Bayes' theorem means that *the probability distribution of ages-at-death (i.e., lifespan) in the target population of interest* must be estimated before $P(a|c)$ can be assessed. This is the prior probability and noted as $f(a)$ in Equation 2.

Note that the denominator of the Bayes' theorem is the integral of the two probabilities, $P(c|a)$ and $f(a)$, across all ages that functions as a constant to give proportionality of the posterior probability, $P(a|c)$.

Although the concept of a prior probability, $f(a)$, may seem to be the most challenging to obtain, it can be derived from by three different sources (Hoppa and Vaupel 2002b). First, a prior probability can be based on the experts subjective judgment when one has ancillary knowledge, and, second, by assuming equal chance of death for all ages. That is 20-year olds are equally likely to die as 5- year olds. Lastly, a prior distribution can be established by using a previously reported age-at-death distribution that is independent from the target sample, and yet large enough and well-approximates the target sample. Specific examples of this third type of informed priors include the 17th century Danish parish records, US homicide records, and Balkan genocide reports on age at missing.

A number of authors have developed age estimation methods based on the framework of the Rostock Manifesto using different age indicators (e.g. first ribs; long-bone lengths (subadults), clavicles, dental structural changes, dental wear, sacral vertebrae, the pubic symphysis, the auricular surface) and diverse populations (e.g. Balkans, the Romano British period infants, modern Americans, Portuguese, South Africans, SE Asians, modern Australians) (DiGangi et al. 2009; Gowland and Chamberlain 2002; Langley-Shirley and Jantz 2010;

Lottering et al. 2013; Lucy et al. 1996; Prince et al. 2008; Ríos et al. 2008; Schmitt 2004), and have shown effectiveness of the Bayesian framework.

Transition Analysis

In line with the spirits of the Rostock Manifesto, Boldsen et al. (2002b) introduced a new method, Transition Analysis (“TA” hereafter), as a comprehensive solution to the issues of age estimation. Although its use in anthropology is relatively new, transition analysis, as a general statistical method, has been widely used in diverse fields of studies, including economics, political science, and engineering. Transition analysis estimates a point of transition from one stage (or state) to next, and can be conducted on any ordinal, categorical variables that satisfy the following assumptions: 1) change from one stage to next is in sequential order (e.g. 1, 2, 3... n), 2) the change is unidirectional (that is, one cannot go back), and 3) the change cannot skip a stage (Boldsen et al. 2002b). In order to run a transition analysis, the researchers tried a number of different options for statistical models. The models include a logit or probit regression, and a backward or forward continuation ratio model to name a few.

Boldsen et al. (2002b)’s method was developed on a little over 180 individuals representing the 19th century African and European Americans of the Terry collection using three age indicators, including the pubic symphysis, the sacroiliac joint, and cranial sutures. In later years, they expanded the reference sample size to over 250 individuals by adding 19th-20th century Portuguese from the Coimbra Collection. The authors adopted a component scoring system for the three key age indicators, and chose skeletal traits that were thought to be the best age estimators. The component scoring system of TA divides age-specific skeletal traits within each age indicator into quantifiable, sequential stages (Table 4 in Chapter 4) so that subjective

visual assessments leading to higher inter-observer error rate can be minimized (Wilson and Algee-Hewitt 2009).

Returning to the three core elements of Bayes' rule, TA calculates parameter values (likelihoods) for ages at transition derived from the Terry and Coimbra collections. Informed priors for TA were established from the 17th century Danish parish records and the 1996 U.S. homicide records by fitting a Gompertz hazard model. Even though the likelihoods were derived from a specific reference sample from a specific time period, the authors argue that TA is free from age mimicry since it can incorporate an appropriate prior, so that the age distribution of the reference samples are not imposed on the target sample age structure. In order to ease its computational complexity, Boldsen et al. developed software specific for TA, ADBOU Transition Analysis Age Estimation 2.1 (Boldsen et al. 2002a). The program provides a posterior probability distribution for each of the three age indicators as well as all three indicators combined.

In their seminal publication, the developers of TA have championed the method as it overcomes the major problems faced in conventional aging estimation methods (Boldsen et al. 2002b; Milner et al. 2008). TA's contributions to age estimation can be summarized as follows: 1) its age estimate output for each individual skeleton is provided as a maximum likelihood estimate (MLE), 50/95/99% credible intervals depends on user's needs, and standard error; 2) it has a reduced age mimicry effect by accommodating different age-at-death distributions appropriate for target samples at interest; 3) it can incorporate within- and between-individual variations by using multiple age indicators to generate final age estimates; 4) it can estimate ages using incomplete skeletons and skeletal elements and; 5) it can estimate ages over 50 years old. Finally, it allows replication due to a composite scoring system; this means that the observer can

examine one skeletal trait on one age indicator at one time instead of evaluating an overall suite of skeletal traits simultaneously, as is the case with phase methods.

Moreover, the outcome age estimates of TA, the posterior probabilities of being in each of ages given skeletal traits, conform to the medico-legal requirement for quantified uncertainty in casework analysis and expert witness testimony in court. For instance, when a biological/forensic anthropologist serves as an expert witness in court, s/he should present to the jury and judge a statistical probability along with the age assessment. This value speaks to the practitioner's confidence in the given estimate being an accurate reflection of “true” age for a particular set of skeletal remains as well as practitioner’s uncertainty in the age estimates (1993, *Daubert v. Merrell Dow Pharmaceuticals* (92-102), 509 U.S. 579) (Committee on Identifying the Needs of the Forensic Sciences Community 2009). Unlike conventional methods, TA provides preferable forms of age estimates in medico-legal settings, and previous publications have demonstrated the potentials of the TA framework to be applicable to such instances (Konigsberg et al. 2008; Steadman et al. 2006).

In 2012, Milner and Boldsen (2012) presented a validation study of TA using modern American samples (n=252). Although the results were not entirely satisfying, the authors addressed several important findings. Overall, at an individual level, TA performed well up to 40 years old, especially when all anatomical elements were combined regardless of using a uniform prior or an informative prior. Thus, TA estimated age distributions that are close to known populations, suggesting that TA captures general mortality patterns well. The least accurate age estimates occurred between 40-70 years, presumably due to a lack of information that the three age indicators provide for those age cohorts. However, error with age estimate did not increase asymptotically. TA assigned individuals beyond 70 years old with better age estimates than those

in the 40s-70s. Beyond the 70s, using an appropriate informative prior can effectively narrow down age intervals of elderly individuals. However, the effects of an informed prior were not significant in young individuals and archaeological populations.

In addition, as anticipated, the pubic symphysis performed the best, although its age intervals became wide in old individuals. On the other hand, the auricular surface and cranial sutures performed poorly in both accuracy and precision. The problem of underestimation was present for 60 years and older for the auricular surface, and over 40 years for cranial sutures. Age intervals of cranial sutures were too wide to provide useful information on age.

Milner and Boldsen's view on what makes a good age estimation technique has evolved since the seminal publication of TA in 2002. In their 2012 study on TA validation, the authors suggest that simple binary variables are as age-informative as variables with multiple stages (Milner and Boldsen 2012). The authors further argue that the traits on the pelvis and the cranium do not provide sufficient information to yield accurate age estimates. They suggest expanding age-indicative traits beyond the pubic symphysis, the auricular surface, and cranial sutures.

The authors are currently in the process of developing a new, experiment-based aging method that uses age-indicative features throughout the entire skeleton. Some of the features include bony lipping at margins of major joints and the vertebral body, ossification of muscle attachment sites, rib cross-section thickness, and depressed parietal bones superior to the temporal lines. Milner and Boldsen had provided a brief preview of this method in their 2012 validation paper that their new method provides superior age estimates when it is combined with the original TA. The efforts to combine the new method to the original TA is currently being

developed by Milner, Boldsen and their colleagues in the U.S. and Denmark (Personal communication 2016).

Chapter 3

The Population History of Southeast and East Asia

In the previous chapter, we have reviewed a number of biological, technical and statistical factors that can lead to disappointing age estimations. Despite such various factors to consider, much focus has been placed on developing ‘population-specific’ age estimation methods under the assumption that skeletons of different populations will age differently. For many years, the field of anthropology, especially the forensic subfield, has encouraged anthropologists to assume that population variation exists in skeletal aging as such that the best practice guidelines for age estimation published by the Scientific Working Group for Forensic Anthropology (SWGANTH) recommend using population-specific (or most inclusive) aging methods whenever possible (SWGANTH 2013).

Ironically, however, in contrast to the fact that there have been many calls for population-specific methods, population variation in aging rates and patterns (if it ever exists) has not been systematically investigated. Thus, few attempts have been made to aggregate those validation studies of different skeletal samples to comprehend what those results actually tell us about population variation in the context of age estimation (Hoppa 2000). And, virtually no emphasis has been placed on developing a more inclusive method that is broadly applicable to diverse (or unknown) populations.

While it is indeed a difficult task to solely evaluate the effect of ancestry in the midst of numerous variables influencing natural skeletal aging process, a carefully designed study with multiple populations with shared population history from the same continent may contribute, to some degree, in teasing apart the impact of ancestry in age estimation. In this sense, Asia serves as a perfect place to understand population-specificity in skeletal aging because the well

documented history on the peopling of Asia and the possible shared ancestral affinities of the continent's peoples gives us a theoretical foundation to design a study that investigates variations in skeletal aging rates and patterns within the same continent. To illustrate the rationale for structuring the current study based on Japanese and Thai skeletal samples, this chapter reviews the population history of Southeast and East Asia, and their shared ancestral history. The chapter also discusses benefits of using these two particular populations for understanding variations in skeletal aging. The chapter closes by identifying two main goals of the current project and by presenting the four hypotheses evaluated in this study.

Why study Asians?

Part of the objective of this project is to compare multiple Asian (sub-) groups and to assess any differences in their skeletal aging rates and patterns. Having multiple reference samples representing a broader ancestral group is advantageous in many ways. First, it provides new knowledge to fill the void of relatively under-represented and under-studied populations for skeletal aging process. This knowledge not only bridges the gap in the academic literature, but also has immediate application to forensic casework in the U.S. According to a brief report of U.S. Census Bureau-Population Division, Asians, as defined as people from the Far East, Southeast Asia, or the Indian subcontinent, including Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam, are the fastest growing ethnic group in the U.S: Between the years of 2000 and 2010, the Asian population in the U.S. increased 43%, and now consists of 6% of the total U.S. population (U.S. Census Bureau-Population Division 2013).

Second, Asian groups make an excellent study sample for investigating population specificity. Tayles and Oxenham (2006, p. 2) eloquently explain that Asia is an ideal locale for “synthesizing research on human biology to address the issues of human evolution, variation and biocultural development,” because it has a unique suite of characteristics not found elsewhere: a very long human settlement, geographic and climate diversity over time and space, and great antiquity of the current rice-based subsistence system. Pechenkina et al. (2013, p. xvii) present a parallel argument and state that this data rich region has a complex population migration and settlement history that has enriched our understanding of the initial peopling of the region. Having in-depth knowledge on the migration routes of the Asian groups and their close biological relationship has its own merits for designing research on skeletal aging because it provides a sound rationale to use Thai and Japanese skeletal samples to evaluate the objectives of this project as outlined in the introduction. Thus it also provides a valid reason to conduct an evaluation of population specificity at multiple levels, including local, sub-continental and continental differences in skeletal aging rates and patterns.

The peopling of Southeast (SE) and East (E) Asia

Migrations of modern humans out of Africa began around 100-150,000 years ago (Mukhopadhyay et al. 2013). A study of mtDNA lineages, or “haplogroups,” suggests that the first founder group arrived in SE Asia about 60,000 years ago via the Indian sub-continental coastline. Some members of this migration wave initially settled in the mainland SE Asia, while others explored further north to E Asia about 25-30,000 years ago (Shi et al. 2008). Numerous studies of population history using genetic data have confirmed these possible migration routes (Ballinger et al. 1992; Jurmain et al. 1994; Lewton 2012; Reich et al. 2010; Reich et al. 2011;

Yao et al. 2002), and there is general agreement on approximately when and to where humankind traveled out of Africa. However, for reconstructing more detailed settlement history at a regional level, such as sub-continent and local levels, there has been no particular consensus. Especially, the peopling of SE and E Asian regions has been at the center of such debates. Note that in this context, Southeast Asia includes modern Myanmar, Thailand, Vietnam, Laos, Cambodia, Malaysia, Singapore, Indonesia, Brunei and the Philippines, whereas East Asia (or sometimes referred to or overlapped with NE Asia) encompasses modern China, Taiwan, Korea, Japan, Mongolia and the Russian Far East (Pietrusewsky 2006).

The most constructive ongoing theoretical debates on the origin of the modern Asian populations are based on two competing models (Oxenham and Tayles 2006; Pechenkina et al. 2013): The first model argues that both SE and E Asian groups are direct descendants of indigenous populations who used to occupy the region since the prehistoric period (the regional continuity model). On the other hand, the second model argues that SE Asians are an admixed population with migrants from NE or E Asia around 4,500 years ago (the two-layer, or replacement model). A wide variety of evidence, including phenotypic and genotypic traits, and archaeological and linguistic records, have been investigated to reconstruct the settlement history of the Asian regions, and different types of evidence support one model over the other. In the field of physical anthropology, dental metric and nonmetric traits and cranial morphometrics that are known to remain consistent over time have been favored as main tools for evaluating the populations' biological relationships (Pietrusewsky 2013).

The first theory, the regional continuity model, argues that the modern SE Asians are evolved from indigenous populations, so-called Australo-Melanesians, who have been occupying the region from the late Holocene onward (Oxenham and Tayles 2006; Pechenkina et al. 2013).

One of the major proponents of the regional continuity model is Christy Turner, who has performed extensive research on dental non-metric traits of Asian skeletal remains (e.g., incisor shoveling, interruption grooves on enamel, number of roots, and extra cusps, to name a few). According to Turner (1989, 1990, 1992a, b), dental characteristics of E and NE Asian groups (“sinodonty”) show more derived features from SE Asian dental characteristics (“sundadonty”). Moreover, the sundadont dental pattern can be observed in sinodont populations, such as the prehistoric Jomon and the modern Ainu in Japan. Therefore, SE Asian groups exhibiting sundadont characteristics are the “ultimate source,” rather than recipients, that subsequently spread E and NE Asia to give rise to sinodonty.

In support of this argument, multivariate craniometric analyses by Hanihara (1993, 1994) and Pietrusewsky (1992, 1994, 1999, 2010, 2013) demonstrate that there is a clear separation between E/NE Asians and SE Asian crania, and a clear connection between mainland and island SE Asian groups. Thus, Chinese, Japanese and Koreans of Northeast Asia share a common recent ancestral origin. Supporting this model, genetic evidence has confirmed that the E/NE and SE Asians are two separate phylogenetic units (Cavalli-Sforza et al. 1994; Omoto 1995; Underhill et al. 2001).

On the other hand, the second (classic) theory, the two-layer or replacement model, states that the present-day SE Asian groups are either mixed with, or replaced by, the NE Asians who migrated into the region during the Neolithic period. Physical anthropologists Matsumura (2001, 2006) and Matsumura and Hudson (2005), favor this model based on morphometric analyses of human skeletal and dental remains. While they agree, in part, with the view that SE Asia was initially inhabited by Australo-Melanesians associated with a Hoabinhian culture during the late Pleistocene and early Holocene periods (8,000-10,000 BP), they hold another view that those

indigenous populations were substantially affected by gene flow from migrants from E Asia (southern China) during the post-Hoabinhian period. As a result, present-day SE Asians are a hybrid group sharing genetic materials of both NE Asian immigrants and the descendant of early indigenous SE Asians (Matsumura 2006; Matsumura and Oxenham 2013). New skeletal remains discovered in Thailand, Malaysia and Vietnam corroborate the two-layer hypothesis (Matsumura 2006; Matsumura and Pookajorn 2005; Matsumura and Zuraina 1995; Matsumura and Zuraina 1999; Oxenham et al. 2008). Also, classic genetic markers and mtDNA analyses demonstrate close biological relationships between Chinese and SE Asian groups (Ballinger et al. 1992; Cavalli-Sforza et al. 1994; Omoto and Saitou 1997; Tan 2001).

Besides biological evidence, archaeological and linguistic studies also substantiate this classic view with the agricultural colonization model (Bellwood 1996; Diamond and Bellwood 2003), which is a compatible theory to the replacement model. The model seeks to explain the prehistory and historical biological relationships of the peoples of E and SE Asia by examining the history of rice domestication and the development of agriculture of the region. According to the agricultural colonization model, NE Asian migration to SE Asia was an “agriculturally-driven demic expansion” (Matsumura 2006, p.60), and is responsible for replacing the earlier indigenous groups of SE Asia (Bellwood 1996; Higham 2001).

In 2006, Tsunehiko Hanihara, who was once a proponent of the regional continuity model, presented results that support both of the models. According to his large-scale analysis based on 34 craniofacial measurements assessing facial flatness of 103 Asian and their neighboring groups (e.g. Australians, Malaysians, and samples of the Arctic, New World and Middle East), there is a clear separation among Australian/Melanesian samples, E/SE Asian samples and NE Asian samples. However, the majority of SE Asians (e.g. Thais) and typical E

Asians (Chinese, Korean, Japanese) showed a close affinity, supporting the two-layer model that indigenous Australo-Melanesians were indeed impacted by N/E Asian migrants from around 5,000 years BP. Nevertheless, the early SE Asians from approx. 5,000 to 10,000 years BP were located between the SE Asian and an Australian/Melanesian sample, while the typical E Asians formed a satellite cluster way from the SE Asian samples, simultaneously supporting the regional continuity model (Hanihara 2006). In addition, although it has been a modern social phenomenon, and therefore not part of any of the proposed models, a more recent migration of southern ethnic Chinese to SE Asia is another non-negligible factor that facilitate gene flows between SE and E Asian groups. According to Vaughn and Morrison (2006), approximately 30 to 40 million ethnic Chinese currently reside in Southeast Asia. Ethnic Chinese are particularly assimilated in Thailand, constituting over 10% of the total Thai populations.

Another theoretical inquiry that is often discussed side-by-side with the origin of SE and E Asians is who were the initial settlers of the Japanese archipelago. A traditional, yet widely-accepted hypothesis is the “dual-structure” model proposed by Kazuro Hanihara (Hanihara 1991). According to this model, the first prehistoric settlers of Japan were Neolithic hunter-gatherers, the Jomon, who are presumably originated from Southeast Asia more than 10,000 years ago. The Neolithic Jomon was later admixed with the second wave of migrants from the E Asian mainland who introduced rice cultivation to Japan, and give rise to the Yayoi people about 2,000–3,000 years ago. This large influx of E Asian migrants is assumed to have entered Japan via the Korean peninsula at two likely entry points, the northern Kyushu and southwestern Honshu of Japan (Pietrusewsky 2013). The prehistoric Jomon has been proposed to be direct contributors of the indigenous Ainu living in Hokkaido, the far northern island of Japan, and also possibly the Ryukyu Islanders in the most southern part of Japan. Because of their geographic

isolation, the Ainu and Ryukyu Islanders are relatively unmixed descendants of the Jomon, hence retaining many physical characteristics of the earlier Jomon people (Hanihara 1991; Turner 1992b), while the rest of the Japanese population has been impacted by gene flow from E Asian migrants. As a result, Japanese populations create a bipolarized, “dual-structured” pattern throughout the archipelago (Hanihara 1991).

Yet, there are two alternative views concerning the first settlers of Japan. Nei (1995) and Omoto and Saitou (1997) argue that the ancient Jomon had a NE Asian origin, whereas Yamaguchi (1992) proposes that the Jomon people rather shared similarities with Upper Palaeolithic hunter-gatherers in Eurasia. Most recently, Pietrusewsky (2013) argues that the ancestral homeland for both ancient and modern Japanese subgroups, including the prehistoric Jomon, the Ainu, and the Ryukyu Islanders, are NE Asia (China and Korea), with Korea being the closest of all. He concludes that a clear dichotomization of SE and E Asian groups implies “long-term *in situ* evolution” in both regions, supporting the regional continuity model. While genetic and morphological evidence has confirmed that the Jomon falls under a broader Asian group (Brace and Hunt 1990; Brace et al. 2001; Horai et al. 1996; Howells 1986; Ishida 1995), the exact origin of the initial settlers of Japan remains unresolved (Hanihara and Ishida 2009).

Why is knowing population history important for developing aging methods?

So far, we have reviewed population history of SE and E Asia and several theories concerning the peopling of the region. Perhaps a broad take home point is that Asian groups are closely related in one way or another, and yet part of them can be grouped separately based on relatively recent (in the evolutionary sense) migration history. Then, how do we integrate all the pieces of information above in a way relevant to modeling age progressive changes in the skeleton?

First, it is worth reiterating that knowing the settlement history of the Asian groups and their close biological relationships provides us a sound reason to use Thai and Japanese skeletal samples to evaluate population variation and specificity in skeletal aging rates and patterns in multiple levels (local, sub-continental and continental differences). Especially, as more Asian reference samples representing local populations become accessible and given the complex migration history of Asia, simply lumping everyone altogether as one single Asian group may not make any logical sense. Hanihara and Ishida (2009) and Relethford (2001) criticize researchers using regional aggregates as the unit of analysis although their criticism is more toward geneticists. According to the authors, such an analytical approach does not elucidate whether the variation observed reflects the diversity *within* local populations, the divergence *among* local populations, or *both*, thus creating false results if there are differences in population substructure within the region.

Second, and most importantly, a review of the population history of Asia and a brief immersion into the population geneticists' mindset encourage us to challenge the idea that each population has unique aging rates and patterns of the skeleton, and instead support the notion of biological uniformitarianism—that biological processes, such as skeletal aging rates and patterns, stay constant throughout time and space (Howell 1976). According to Howell (1976), who first defined the biological uniformitarianism by borrowing the term from the field of geology:

“A uniformitarian position in paleodemography implies that the human animal has not basically changed in its direct biological response to the environment in processes of ovulation, spermatogenesis, length of pregnancy, degree of helplessness of the young and rates of maturation and senility over time. This does not imply that humans have not changed in the rates of performance of these processes, but only that the processes still respond in the same way to variations in environment; including the cultural and technological aspects of human society as part of the external environment. Another way of stating this is to assert that the demographically relevant biological processes of our species are constant in our genetic composition, subject only to variation in response to environmental forces, and that the species has not undergone any significant intra-species evolution since its first appearance as *Homo sapiens*” (p. 60).

Both skeletal morphometrics and genetic information indicate: (1) a common basal origin from Africa and subsequent migration out of Africa to the rest of the major geographic regions, including Europe, Asia, Polynesia, and Americas; and (2) our adaptation through natural selection with close relation to climate and subsistence strategy (Long 2004; Long and Kittles 2009; Relethford et al. 1994; Relethford 2004a, b, 2009). In other words, as indicated by our biological information, human biological variation is indeed geographically structured, but also structured in a way that certain populations can be grouped closer together than others. These findings further suggest a possibility that, in partial support of the biological uniformitarian assumption, applying an osteological standard developed on a certain reference population *may* work better for another target population with a closer relationship with the reference group than another population both geographically and historically distanced.

The traditional idea has been that different parts of the skeleton render different information on the history of humanity. Specifically, selectively neutral traits that do not affect fitness or survival enable us to correctly trace back our shared population history (the neutral theory of evolution) without being misled by biological variation caused by natural selection. At the same time, however, highly selective traits reflecting our adaptation to extreme environment have provided information on individuals' "biocultural" affinity to a specific population (i.e. ancestry estimation). A great example of this concept is the cranium as the studies of SE and E Asian population history have shown above. While craniomorphometric variation can be considered selectively neutral, certain cranial traits are affected by selection (Relethford 2002). Roseman (2004) argues that cranial measurements reflecting thermoregulatory capacity in colder environment, such as the breadth of the skull, and other measures supporting the thermoregulatory breathing hypothesis, such as nasal height and nasion-prosthion height,

deviated from neutral expectations, while measurements reflecting the positioning of the face were in accordance of population history and structure. In line with this argument, Smith (2009) concludes that the morphology of the temporal bone, upper face, basicranium, and entire cranium consistently reflect genetic relationships in humans, whereas the morphology of the mandible, upper jaw, and cranial vault do not, given those cranial sites are particularly plastic in nature due to adaptive responses to mastication and/or remodeling of alveolar processes (Baumhammers et al. 1965; TenCate and Mills 1972).

Interestingly, von Cramon-Taubadel (2014) presents striking findings that even cranial elements that are related to functional and climatic adaptation still convey phylogenetically insightful information. Thus, the results have shown that specific function or ossification origin (cartilaginous vs. fibrous model) does not influence or significantly obscure underlying population history and structure, and therefore those functional cranial elements can still meet the neutral expectation (though the neutrality can be weak). However, a more recent, comprehensive craniometric study by Roseman (2016) has demonstrated that the cranium offers a little information to reconstruct recent human evolution, and the cranial variation caused by relatively recent population history and adaption overrides the neutral expectation.

Betti et al. (2013) study has shown that the neutral model is also applicable to postcranial elements, such as the pelvis, which has been long considered to be highly selective as it has high correlations to the body size and proportion adapted to extreme climates. Recently, Savell et al. (2016) has argued that between-group differences in the limb bones, including tibia, radius, femoral head and biliar breadth could be, in part, explained as signs of directional selection, but in fact the femur was subject to minimal or no directional selection, and between-group variation in the humerus was rather a result of correlated responses to natural selection acting on other

traits or skeletal elements. Additionally, a study of Anderson and Trinkaus (1998) demonstrate that there are no significant inter-population differences in, for this case, human femoral neck-shaft angles, and no correlations with latitude. Rather than innate ancestral differences, a significant increase in mean neck-shaft angles across populations was more associated with an increased sedentary life due to mechanization, further indicating developmental plasticity with respect to habitual load levels during ontogeny of the pelvic region.

Now returning to the current project, recall that the ultimate goal of this project is to investigate the level of population-specificity in skeletal aging rates and patterns, and to evaluate if two geographically and historically close Asian populations can be pooled together to develop a more collective age estimation methods. To answer these questions, the project uses three traditional age indicators, including the sutures of the cranium, and two pelvic joints, namely the pubic symphysis and the sacroiliac joint. Also recall that these are the same skeletal elements that have been used to reconstruct our common ancestral history, migration routes and climatic adaptation. This fact then makes us pause and question: Will age-related traits (which is more nuanced than the general shape of the bones) on these skeletal elements be also subject to microevolutionary forces associated with climatic adaptation and environmental plasticity? Looking at a bigger picture by combining all the aforementioned studies, it might not be entirely unreasonable to have neutral expectations of skeletal aging as well. A counterargument may be that it cannot be interpreted in such a way since the current project's methodological approach is macromorphoscopic, while the studies evaluating the neutral theory of expectation are based on metric analysis. However, it is important to note that metric measurements are, after all, a reflection of morphology. Moreover, age estimation is already, in some sense, based on a similar concept of neutral expectation, as anthropologists often work under an assumption that the

sequence of skeletal developmental and aging “patterns” is the same across all populations. What has been at the center of the argument concerning population-specificity and variation in skeletal aging is the “rate” or “speed” of transition of age-related skeletal traits from one stage to next.

To summarize this chapter, it is timely in the fields of skeletal biology and forensic anthropology to stop uncritically equivocating to develop population-specific aging methods, and to think beyond the taxonomic expectation that people from different geographical regions should show differential skeletal aging processes just as we would expect from skeletal morphometric traits used for ancestry estimation. If the studies of population and quantitative genetics inform us that we have maintained signals of our shared ancestral history for hundreds and thousands of years throughout the skeleton, then does it not give us a valid justification to think that we might be able to develop one single universal age estimation method, at least for the Asian populations, as opposed to multiple population-specific methods?

Hypotheses

The previous discussion on the population history of SE and E Asia leads us to identify the two main goals of the current study. One is to quantify the accuracy and reliability of some current age estimation methods (TA and conventional methods) used for estimating the skeletal aging within Asian populations (or any other non-North American populations). The other is to help understand population-specificity in skeletal aging rates and patterns between populations from the same continent. In order to achieve these two goals, the project is testing four hypotheses, as presented below, that begins with validation studies of TA and conventional methods, with much emphasis on the performance of TA, on four Asian skeletal collections with fully documented demographic information. To help understand population-specificity in

skeletal aging rates and pattern between populations within the same continent, four Asian skeletal collections of Japanese and Thais, representing part of E and SE Asia were used.

Hypothesis (1) Based on the findings of previous studies that have shown population-specificity in skeletal aging (Katz and Suchey 1989; Lottering et al. 2013; Schmitt 2004), it is hypothesized *that the current transition analysis method will yield age estimates high in error and bias for Japanese and Thai skeletal samples because the method was originally developed on non-Asian populations.* To obtain age estimates for Asian populations, the TA program, ADBOU Age Estimation 2.1, is used.

Contrary to previous studies, only a few argue for the non-existence (or non-importance) of population-specific aging (Kimmerle et al. 2008; Konigsberg et al. 2008). Moreover, while published validation studies of TA are remarkably few in number (Table 2), results of those studies were contradictory. Bethard (2005) and Milner and Boldsen (2012) argued that the performance of TA was not entirely unsatisfactory for age estimation of modern Americans, while Wilson and Algee-Hewitt (2009) and Algee-Hewitt and Wilson-Taylor (2011) stated that TA provides promising age estimates for modern Americans and Southwest Hispanic populations. This first hypothesis will eventually help shed light on the effectiveness of both the pre-existing transition ages, derived from the original reference samples, Americans and Portuguese, and informed priors fixed in ADBOU for achieving accurate and reliable TA age estimates of Asian samples. Currently, ADBOU offers two informed priors including the 1996 U.S. homicide cases (Peters et al. 1998) and the 17th century Danish parish death record (Johansen 1998). Thus, testing this hypothesis will quantify the magnitude of error and bias of

TA age estimates for Asian samples that will further serve as important baseline to compare age estimates derived from Asian-specific parameters created in the next hypotheses.

Table 2. Sample size and composition for TA validation studies

Validation Study	Period (collection name)	Sample size
Boldsen et al. (2002b)	A 19 th cen. U.S. sample (the Terry Collection)	84
	A 17 th century Danish sample (Arhus Cathedral, a case study)	1
Bethard (2005)	A modern U.S. sample (the William Bass (WB) Donated Collection)	225
Wilson and Algee-Hewitt (2009)	A modern U.S. sample (the William Bass Forensic Collection)	58
Algee-Hewitt and Wilson-Taylor (2011)	Modern Southwestern Hispanic and U.S. samples (the Pima County Collection, the William Bass donated and Forensic Collections)	63
Milner and Boldsen (2012)	A modern U.S. Sample (the William Bass Donated Collection and Mercyhurst Forensic Collection)	252

Hypothesis (2) Given that results of the previous studies have shown greater variability of female skeletons than males due to biomechanical stress related to childbearing and child birth (Kemkes-Grottenthaler 2002; Suchey and Katz 1998), it is hypothesized that *there will be a certain degree of sex differences in age progressive changes between Japanese (or Thai) male skeletons and Japanese (or Thai) female skeletons*. Moreover, although maturation and degeneration are not synonymous, it is now well reported that the female skeletons tend to mature faster than male skeletons (Algee-Hewitt 2012). If the fast maturation rate in female skeletons continuously leads to faster degeneration, and the different function of the female pelvis also contributes to increasing variability in females, then the second hypothesis is expected to be supported and therefore, sex-specific age estimation models will be necessary for the two Asian populations.

Hypothesis (3) In contrast to the findings of the previous studies that have shown population-specificity in skeletal aging (Katz and Suchey 1989; Lottering et al. 2013; Schmitt 2004), it is hypothesized that *an age estimation model developed on Japanese will not necessarily produce inaccurate age estimates for Thais, and vice versa*. The two rationales for building this hypothesis come from the closely shared population history of Japanese and Thai populations, and neutral expectations of age indicators used in this study—the two pelvic joints’ surfaces and cranial sutures that are more nuanced features than the overall morphology of the pelvis and cranium, and therefore, possibly not dependent on the environment and natural selection—as already reviewed in this chapter. This hypothesis ultimately evaluates the validity of the uniformitarian assumption (Howell 1976) that population variation in aging rates stays constant across time and space. The hypothesis also addresses whether or not the Japanese and Thai samples can be pooled together to collectively develop a single age estimation model for Asians.

Hypothesis (4) Further building on the two rationales of Hypothesis (3) for expecting a lack of population specific skeletal aging between two Asian groups, it is hypothesized that *age estimation models based on pooled Japanese and Thai samples will perform better than Japanese- and Thai-specific models and ADBOU or conventional age estimation methods by generating more accurate and unbiased age estimates for the Japanese and Thai samples*.

Theoretically, when an age estimation method was established based on the Bayesian framework by incorporating adequate reference samples (likelihood functions) and appropriate prior information on an age-at-death distribution of the population at interest (an informed prior), age estimates should be accurate and reliable with no issue of age mimicry. Previous studies have shown promising results of such a Bayesian approach with improved accuracy and reliability in

age estimates (DiGangi et al. 2009; Kimmerle et al. 2008; Konigsberg 2015b; Konigsberg and Frankenberg 2013; Prince et al. 2008).

Chapter 4

Methods

The methods described in this chapter are chosen to test the hypotheses described in Chapter 3. The purpose of the study is to quantify the magnitude of error and bias of TA and three conventional methods when they are applied to Asian samples, and to investigate whether each Asian population needs an aging method exclusive to the group, or if two different populations can be pooled together to develop one aging method. Therefore, this chapter gives an overview of the Asian samples and the methods designed to uncover error and bias. The chapter illustrates multiple steps to model age progressive changes of the Asian samples, and describes how age estimates are retrieved from the skeletal data and mortality data.

Skeletal samples

The current project uses four different skeletal collections from Thailand and Japan. These collections serve as an ideal sample to test the four hypotheses outlined in the previous chapter because they represent part of Southeast and East Asians who share a close population history. This allows me to explore the question whether population specificity in skeletal aging rates and patterns exists at a continental level or a sub-continental level. Thus, these four relatively understudied collections will contribute to filling the void in the literature related to the skeletal aging of Asian populations. Also, the four collections will quantify error and bias associated with using Boldsen et al.'s (2002b) TA for Asian populations, as the method has not yet been tested on various populations as conventional methods (Table 2 in Chapter 3).

The four documented skeletal collections (total number of skeletons =1,613) used for this project are currently housed in the following institutions: The Department of Environmental

Medicine, Chiba University, Chiba, Japan; The Department of Anatomy, the Jikei University School of Medicine, Tokyo, Japan; The Department of Anatomy, Khon Kaen University, Khon Kaen, Thailand; and The Forensic Osteological Research Center (FORC), Chiang Mai University, Chiang Mai, Thailand.

Chiba and Jikei collections in Japan represent Japanese who were born in the early 1900s and the late 1800s and died during the 20th century at university hospitals. The Jikei Collection also includes cases from medical examiners' office. The Chiang Mai FORC and Khon Kaen collections in Thailand consist of late 20th-21st century modern Thais procured from a body donation program (Table 3). All individuals examined for this study were fully documented with relevant demographic information. Given that the proposed project focuses on aging of adult skeletons, sub-adult skeletal remains were excluded. As a result, the youngest chronological age in the current study samples is 14 years, while the oldest is 97 years.

Table 3. Sample composition of the four collections used in this study

Location	Collection	Female	Male	Total
Japan	Chiba	55	145	200
	Jikei	69	214	283
Thailand	Khon Kaen	240	460	700
	FORC Chiang Mai	154	276	430
Total		518	1,095	1,613

Data collection

Scoring for conventional age estimation methods

Three conventional age-at-death estimation methods were applied to subsets of the Asian samples (total n=259) in order to test for its accuracy and reliability in comparison to TA. Three classic age estimation methods were applied, including Lovejoy et al. (1985a) for the sacroiliac

joint, Suchey et al. (1986) for the pubic symphysis, and Meindl and Lovejoy (1985) for the ectocranial suture closure. Lovejoy et al.'s (1985a) method evaluates the sacroiliac joint based on seven broad criteria, including changes in the apex, superior and inferior demiface topography, presence or absence of billowing, fine vs. coarse granularity, density of the bone, micro- or macro-porosity and bony growths (exostoses) on the retroauricular aspect. Based on a combination of characteristics, each bone is assigned to one of the eight phases. For Suchey-Brooks' method, the pubic symphysis is assigned to one of the six phases based on an extent of ridges and furrows, development and degeneration of ventral and dorsal aspects of the symphyseal face and porosity. Lastly, for the method Meindl and Lovejoy (1985), five vault sites, including midlambdoid, lambda, obelion, anterior sagittal, and bregma sutures, and seven lateral-anterior sites, anterior sagittal, bregma, pterion, midcoronal, sphenofrontal, inferior and superior sphenotemporal sutures are scored based on an ordinal scale ranging from 0 (open) to 3 (obliterated). Once every site is scored, scores are summed to assign each individual into a relevant phase. The vault sites have six different phases, while the lateral-anterior sites have seven phases.

Transition Analysis scoring

For TA, 742 individuals with complete demographic information were examined based on Milner and Boldsen's (2013) most recent scoring protocol. Age indicators used for TA scoring parallel those of the conventional methods, including the pubic symphysis, cranial sutures, and the sacroiliac joint of the ilium including the retroauricular surface. In the TA scoring system, each of the three age indicators is further divided into multiple components (Table 4). Each of the components has multiple sequential stages on an ordinal scale. When all three age indicators

are available for a complete skeleton, 19 different components (five cranial suture sites, five pubic symphyseal components, and nine auricular surface components) with 89 different trait expressions can be evaluated for age estimation from one side. Because TA is based on a component-scoring system, fragmentary bone elements could still be scored not only when one or two of the three age indicators were available for an individual, but also if an age indicator has few components to score for age estimation (missing data). Additionally, following the recommendation of Buikstra and Ubelaker (1994) for evaluating observer's internal consistency in scoring, approximately 10% of the total sample size ($n=82$) was re-scored for the second time at least two weeks after the initial scoring.

When applying TA and conventional methods, the observer was blind to the known age of each of the individuals until data collection was complete to simulate actual forensic cases or bioarchaeological analyses. In order to minimize biased assessment, each of the three age indicators was scored separately for both TA and conventional methods scoring, thus re-visiting each of the skeletons three times in total for independent scoring. This approach helps prevent an observer from making an "age-informed scoring" on the second age indicator or a component based on the first age indicator or a component that was previously scored. The first round of scoring, which only involved TA traits, only included cranial sutures scoring, followed by the pubic symphysis, and the sacroiliac joint. Scores for conventional methods from the same three age indicators were collected at least a week after TA scoring was conducted. Individuals with pathological conditions (e.g. a fused vertebral column, sacroiliac fusion) were not necessarily excluded to incorporate as much variation in age-related skeletal expression as possible. Any available side of the bones was scored and age estimates were retrieved based on either side of

Table 4. Abbreviated scoring format for TA (Milner and Boldsen 2013)

(1) Pubic Symphysis

Score	Symphyseal Relief	Symphyseal Texture	Superior Apex	Ventral Symphyseal Margin	Dorsal Symphyseal Margin
1	Sharp billowing	Smooth	No protuberance	Serrated	Serrated
2	Soft, deep billowing	Coarse grained	Early protuberance	Beveling	Flattening incomplete
3	Soft, shallow billowing	Microporosity	Late protuberance	Rampart formation	Flattening complete
4	Residual billowing	Macroporosity	Integrated	Rampart completion I	Rim
5	Flat	N/A	N/A	Rampart completion II	Breakdown
6	Irregular	N/A	N/A	Rim	N/A
7	N/A	N/A	N/A	Breakdown	N/A

(2) Sacroiliac Joint

Score	Superior & Inferior Demiface Topography	Superior & Inferior Surface Morphology	Inferior Surface Texture	Superior & Inferior Posterior Iliac Exostoses	Posterior Spicules
1	Undulating	>2/3 covered by billows	Smooth	Smooth	Smooth
2	Median elevation	1/3-2/3 covered by billows	Microporosity	Round bony elevations	Rounded spicules
3	Flat to irregular	Flat (no billows)	Macroporosity	Pointed exostoses	Pointed spicules
4	N/A	Bumps	N/A	Jagged exostoses	N/A
5	N/A	N/A	N/A	Touching exostoses	N/A
6	N/A	N/A	N/A	Fusion	N/A

(3) Cranial Sutures

Score	Coronal Pterica	Sagittal Obelica	Lambdoidal Asterica	Zygomaticomaxillary	Interpalatine
1	Open	Open	Open	Open	Open
2	Juxtaposed	Juxtaposed	Juxtaposed	Juxtaposed	N/A
3	Partially obliterated	Partially obliterated	Partially obliterated	Partially obliterated	Partially obliterated
4	Punctuated	Punctuated	Punctuated	Punctuated	Punctuated
5	Obliterated	Obliterated	Obliterated	Obliterated	Obliterated

the age indicator. When both the left and right sides of the age indicator were present, both sides were scored.

Data analysis

Intra-observer error test

Prior to data analysis, it is essential to test observer consistency in following age estimation protocols. Intra-observer error tests for conventional methods were only performed on TA scores, excluding scores of the conventional methods as the main focus of the current study was modeling age-related changes of Asian populations using the TA method. Thus, given that TA evaluates the exact same age indicators as conventional methods, consistency of the observer in applying conventional methods can also be inferred from the observer error test of TA.

To evaluate the intra-observer error rate, the raw scores of TA were obtained from two independent trials of scoring on randomly selected 82 individuals from a total of 742 Japanese and Thai individuals. Agreement between the first and second scores was assessed by using a series of weighted-Cohen's kappa coefficient (Cohen 1960) on each of the skeletal traits scored. An inter-rater reliability test package, 'irr' (Gamer et al. 2012), specifically designed for R, was consulted to perform the analysis. Cohen's kappa (κ) not only evaluates internal consistency of the observer, but also provides information on which specific skeletal feature is more likely to be inconsistent. Landis and Koch's (1977) recommended kappa values were consulted to assess strength of agreement between the first and second scoring: $k = 0$ for no agreement, $k = 0.41-0.60$ for moderate agreement, $k = 0.61-0.80$ for substantial agreement, and $k \geq 0.81$ for almost perfect agreement (Table 5).

Table 5. Kappa statistics and strength of agreement (Landis and Koch, 1977, p. 165)

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Hypothesis 1. Testing accuracy and reliability of ADBOU Age Estimation 2.1

In order to evaluate the performance of Boldsen et al.'s (2002b) original method on the Asian samples, TA age estimates were calculated using ADBOU Age Estimation 2.1 (Boldsen et al. 2002a). ADBOU generates five different posterior density distributions showing different probabilities of being at a particular age given the observed skeletal traits along with a point estimate (a maximum likelihood estimate, or MLE), and an estimated age range (a credible set). The five posterior density distributions of ADBOU include estimates of the three age indicators based on a uniform prior, and two distributions of all three age indicators combined using informed and uniform priors. For the Asian samples, final age estimates from each age indicator and age estimates of all three indicators combined were generated based on a uniform prior. In addition, with an absence of an adequate informed prior in ADBOU for Asian populations, the 17th century Danish prior was used to obtain age estimates of all three age indicators combined as it represents older individuals better than the homicide prior.

The performance of ADBOU was evaluated in two ways. First, point estimates (MLEs) of TA were compared to documented chronological ages of the Japanese and Thai skeletal samples. More specifically, the root-mean square error (RMSE) was calculated using documented chronological ages and estimated ages (Equation 3).

$$RMSE = \sqrt{\frac{\sum (Estimated\ age - documented\ age)^2}{n}} \quad (3)$$

RMSE is a square rooted value of a sum of squared differences between estimated and documented ages divided by the sample size. As a result, RMSE values can easily be interpreted in the original scale (year). Second, bias of ADBOU was measured using the difference between estimated age and documented age, so that any over- and underestimation trend can be examined. For the validation of ADBOU, a total of 742 individuals were used. However, a sample size varies for each of the three individual age indicators, as not were all three elements always available in the Asian samples.

Modeling age-progressive changes in the Asian samples for Hypotheses 2-4

The statistical testing of Hypotheses 2-4 closely follows protocols outlined in Konigsberg et al. (2008), Konigsberg and Frankenberg (2013), Konigsberg (2015b), and Konigsberg et al. (2016). This part of the analysis performs two series of Markov chain Monte Carlo (MCMC) methods, which are a group of algorithms for randomly sampling parameters from a desired probability distribution to fit a statistical model. First, MCMC methods were used to obtain parameters (slopes and intercepts) of the multivariate ordered probit regression (a.k.a. latent trait analysis) using ordinal scores of the skeletal traits and log-transformed documented known ages. The second MCMC sampling methods generated final age estimates for the Asian samples by integrating the probit parameters and an informed prior.

Hypotheses 2-4 involve complex computational steps. Therefore, each of the steps is discussed in greater detail below with justifications why TA and conventional scores are analyzed based on this particular methodological approach (subheadings (a)-(e)). All of the R

scripts used in this study are written by Konigsberg and freely available online (Konigsberg 2015a). Because the current R scripts for the multivariate regression do not allow missing values in the dataset, Asian individuals with missing traits are omitted. As a result, 419 individuals, including 295 males and 124 females were used for this part of the analysis. An exact sample distribution and descriptive statistics are presented in Figure 1 and Table 6.

(a) Missing stages and stage adjustments

Prior to applying the probit regression and the MCMC methods to the Asian data, slight adjustments had to be made to the scoring system in order to facilitate the data analysis for TA trait stages with a sample size of zero or too few individuals. If the first stage is absent, the scoring system simply starts from the second stage. Stages with too few individuals were collapsed with either the next lower or next higher stage. TA traits that were subject to stage collapsing or adjustments include two traits of the pubic symphysis and five traits of the sacroiliac joint: the symphyseal relief and texture, superior, apical and inferior morphology, and the posterior superior and inferior exostoses.

For the pubic symphyseal relief, the stage system starts from the second stage (round, deep billows) given absence of the first stage (sharp, deep billows), resulting in a five-stage system instead of the original six-stage system. Likewise, the superior apex of the pubic symphysis starts from Stage 2 (early protuberance) due to absence of Stage 1 (no protuberance) resulting the trait to be evaluated on a 3-stage system instead of 4.

For the 3 traits of the sacroiliac joint, superior, apical and inferior demiface morphology, Stage 1 and Stage 2 were collapsed due to a small number of individuals in Stage 1. Similarly, Stage 4 (Jagged exostoses), Stage 5 (Touching exostoses), and stage 6 (Sacroiliac fusion) of the

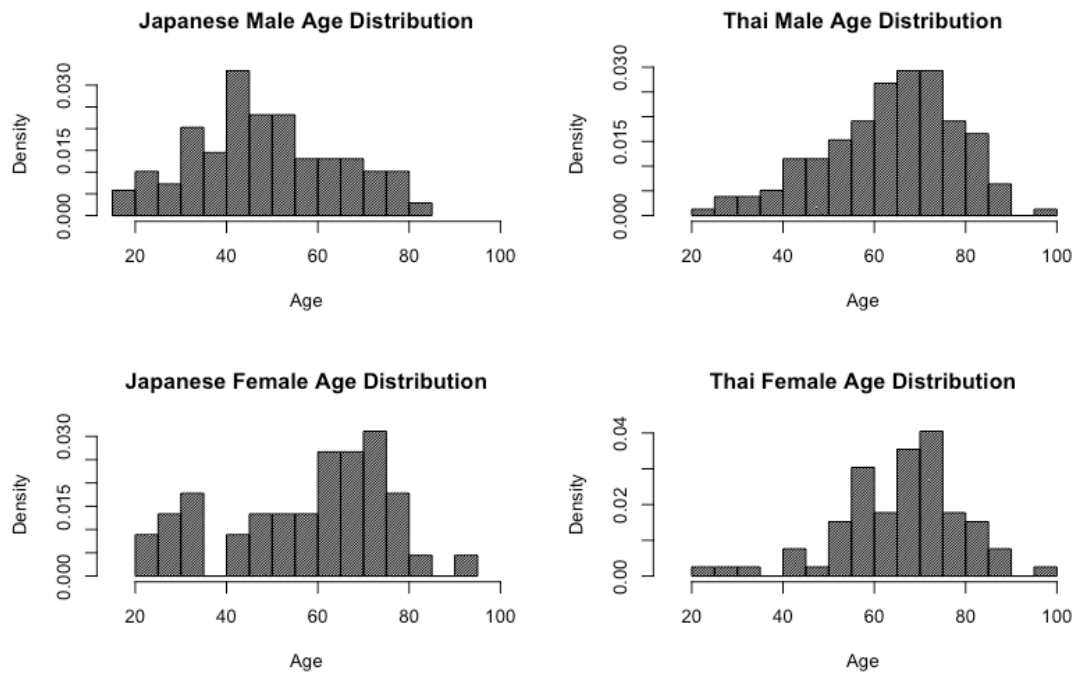


Figure 1. Sample distributions of Japanese and Thai males and females

Table 6. Descriptive statistics for the Asian samples (n=419)

	n	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Japanese Males	138	17	38	48	49.07	59.75	85
Japanese Females	45	23	48	63	58.07	72	95
Thai Males	157	22	54	66	63.68	73	96
Thai Females	79	20	58.5	69	66.46	74.5	97
Pooled Males	295	17	45	58	56.85	70	96
Pooled Females	124	20	56	68	63.41	73.25	97

posterior superior and inferior exostoses were collapsed due to a lack of samples for stages 4 and 6.

(b) Multivariate ordered probit regression

To fit the multivariate ordered probit regression model, two variables were used: the ordered stages describing age-related changes of the Asian samples, and (log-transformed) known chronological ages. In order to avoid the issue of imposing a reference sample age structure to age estimates, ordinal stages were treated as a dependent variable and known ages were treated as an independent variable in the probit regression (Konigsberg 2015b; Konigsberg et al. 2008).

Probit regression offers a number of advantages over other regression models. When a dependent variable is in an ordinal scale—that is, neither continuous nor normal—linear regression models cannot be applied, despite having an independent variable on a continuous scale (c.f. known age), due to a non-linear relationship between the dependent variable and the independent variable. One solution is to use the generalized linear model (GLM). Scholars have shown different kinds of GLM regression to model an ordinal-scale dependent variable with a continuous independent variable. Two most popular models have been logit and probit regression. As briefly mentioned in Chapter 2, Boldsen et al. (2002b) employ a continuation ratio logit regression for ADBOU Age Estimation. Konigsberg and Herrmann (2002); Ríos et al. (2008); and Rissech et al. (2012) use an unrestricted cumulative probit regression. In addition, Samworth and Gowland (2007) developed an infant age estimation method based on shifted exponential data.

Despite the multiple options for modeling age-related changes, Konigsberg has been a long-term advocate of using the probit regression (Kimmerle et al. 2008; Konigsberg and

Frankenberg 2013; Konigsberg et al. 2008; Prince et al. 2008). The probit regression can be extended to the “ordered” probit regression when a dependent variable has more than two ordered outcomes. The probit model essentially “links” an ordinal, categorical response variable to a continuous predictor variable, such that the model is sometimes referred to as “a link function.” According to Konigsberg et al. (2008), the probit regression, especially with chronological age on a natural-log scale, is not only an appropriate approach for the type of data analyzed for age estimation, but also enables researchers to present the statistics in a simple tabular form and to use the parameters for additional analysis. Lynch (2007) also advocates using the probit regression over the logit regression, because the probit regression is linear in standard normal unit (z), parameter interpretation of the probit regression is easier than the logit regression (Lynch 2007, p.197).

The probit regression assumes that morphological changes in an age indicator can be thought of as a continuous process on an underlying, invisible metric scale. This continuous scale of the aging process has thresholds that differentiate one stage to another that eventually enable us to divide lifelong changes of the age indicator into multiple ordinal discrete stages. In other words, a threshold is a delimiting point where one stage moves to the next. Because of this, a trait with n stages will always have $n-1$ thresholds. It is further assumed that the ordinal scores are generated “randomly” from the underlying metric variable, which is normally distributed (Kruschke 2014, p. 672). A more sophisticated way of treating this concept is to consider the underlying metric values of the age indicator as a “latent variable” and generate the ordinal stages with a thresholded cumulative-normal model (Kruschke 2014, p. 672). In latent trait analysis, the probit regression model is expressed as a linear model for a continuously distributed unobserved trait (response variable, denoted as y^*) (Lynch 2007, p. 196). This latent trait

approach becomes useful for setting up the Markov Chain Monte Carlo (MCMC) methods, the methods that the current analysis is also employing.

For the analysis of Hypotheses 2-4, an R script of the multivariate probit regression and additional necessary accessory functions written by Konigsberg (2015b) were used. Latent trait slopes and intercepts (thresholds) were obtained from the multivariate ordered probit regression using the MCMC Gibbs sampler (1000 iterations).

Outcome values of the probit regression include a constant for each predictor variable (known age) and coefficients for slope and intercepts (or threshold). The slope and intercepts are used to calculate a mean and standard deviation of (log-transformed) age at transition from one skeletal trait stage to a next higher stage such that:

$$\text{Mean age} = \frac{\text{intercept}}{\text{slope}} \quad (4)$$

$$\text{Standard Deviation} = \frac{1}{\text{slope}} \quad (5)$$

(c) Prior selections

In order to make the probit model fully Bayesian, priors on the regression parameters in the model need to be specified (Lynch 2007, p.197). Prior distributions can be vague, such as a uniform prior, or specific, such as an informed prior. To establish an informed prior appropriate for the Japanese samples, and to assess effects of different priors to final age estimates, two independent informed priors were created. One is based on large-scale Japanese mortality data ranging from 1947 up to 1995 (Japanese Mortality Database 2014). This wide range mortality database was considered adequate for the Japanese samples given that the Chiba and Jikei collections were procured from early 1900s up to 1995. The second informed prior is based on

documented chronological ages of the Chiba and Jikei collections. Due to absence of a published mortality database for Thais, informed priors for the Thai samples were established by fitting the Gompertz hazard function to the documented chronological ages of the two Thai collections. Konigsberg et al. (2008) once criticized using reference samples to obtain both likelihood and prior probabilities. However, when a reasonable informed prior is not available for the target sample to calculate age distribution, it is possible to use the reference sample's age distribution as a prior (Konigsberg et al. 2008).

(d) Using a simulation approach to calculate posterior probabilities

In Bayesian inference, two conditions are required for mathematical tractability: 1) the likelihood and the prior should share the same form of function, so that the posterior can also be expressed with a similar form to the prior. This type of prior is called “a conjugate prior.” 2) The denominator of Equation 2 in Chapter 2 should be solvable analytically (Kruschke 2014). When the likelihood function and prior distribution have mathematical forms, two mathematical forms can be multiplied together in the numerator part of the Bayesian theorem (Equation 2 in Chapter 2). However, in the case of a multi-parameter probit regression, the computational process to obtain integrals for the posterior density becomes extremely complicated. While a known likelihood form and use of conjugate prior enable practitioners to numerically calculate posterior densities, when forms are unknown, one needs to resort to a computer-based simulation approach. This is called the Markov chain Monte Carlo (MCMC) methods. The idea is to accurately approximate the posterior distribution by a very large representative random sample of parameter values drawn from the posterior distribution. As argued by Kruschke et al. (2012), recent advances in algorithms and software have made it possible for a researcher merely to

specify the form of the likelihood function and prior distribution, and the software is able to apply MCMC methods to obtain the posterior probability distribution

To generate final age estimates, the MCMC random-walk Metropolis-Hastings sampler was performed with a “thinning interval” of five iterations, and the first 200 burn-in samples were dropped from the saved 1200 iterations for each individual (Konigsberg 2015b, p.374). Using these sampled “candidate” ages, a mode (recommended by Konigsberg, personal communication 2016) and 95% highest posterior density (HPD) regions were obtained as a point estimate and a range estimate, respectively. Instead of assuming conditional independence of each of the skeletal traits, Konigsberg (2015b, p.375) suggests accounting for non-zero residual correlation between skeletal traits after partialling out the effect of age. Therefore, a residual correlation matrix of all traits that was obtained from Step (b) was integrated into the analysis for obtaining final age estimates. All of the analysis was performed in software environment R (<https://www.r-project.org>). Besides the R functions of the MCMC sampling methods written by Lyle Konigsberg, two additional R packages, TeachingDemos and Modeest, were used to attain the HPDs and point estimates.

(e) The log-normality assumption check

Since the assumption of a log-normal distribution for estimated ages was made for the analysis, checking reasonableness of the assumption is in order. This can be done in two different ways: (1) Shackelford et al. (2012) recommends a visual inspection of transition distributions as well as evaluating values for log-mean ages at transition. 2) Recently, Konigsberg (2015b) and Konigsberg et al. (2016) suggested the Lagrange multiplier test of goodness-of-fit for the ordered probit regression (Glewwe 1997; Johnson 1996a; Weiss 1997) that conveniently offers both the

normality check and the model fit test. When the Lagrange multiplier test indicates that scores of a certain trait deviate from normality, two or more stages of that trait can be collapsed to optimize the fit of the model, instead of resorting to an alternative parametric model or non-parametric model. R functions to carry out this test are available at Konigsberg's website (Konigsberg 2015a).

However, the currently available R script for the Lagrange multiplier test only concerns a univariate probit regression. Therefore, even though a Lagrange multiplier test indicates an inadequate fit of a univariate probit regression model to the data at hand, it does not equally mean a multivariate probit regression is also inadequate for the data. Regardless, Konigsberg et al. (2016) argues that violation of log-normality on a univariate scale further suggests that normality in multivariate scale is not warranted. Preliminarily, the Lagrange multiplier test was conducted on the sex- and population-specific Asian samples (not presented), and the test statistics indicated a number of traits need stage collapsing. However, if stage collapsing is followed, stage-by-stage comparisons for sex differences and between-group variation in aging rates become extremely complicated. Therefore, for the analysis of sex and population difference in aging rates, the suggestion of Shackelford et al. (2012) was adopted to evaluate male-female differences and population differences in aging rates. For pooled-sample analysis, however, the Lagrange multiplier test and the stage collapsing procedure were conducted.

Now that all the analytical steps are reviewed in subsections (a)-(e), we can return to the remaining hypotheses. The next section explains how these steps apply to test each of the three hypotheses.

Hypothesis 2. Evaluating sex differences in aging rates

In order to evaluate sex differences in speed of aging for each skeletal trait, mean ages at transition from one stage to the next higher stage were examined. Mean ages at transition and standard deviation (SD) can be retrieved using slopes and intercepts of the multivariate ordered probit analysis as well as Equations 4 and 5. The probit analysis provides sex- and population-specific slopes and intercepts for each of the 33 traits for Japanese males and Thai females. The log-mean age at transition and SD calculated from the probit parameters were further used to recover a mode, median and mean in the original scale (year). Due to the asymmetric nature of the log-normal distribution, the mode, median and mean values are not always the same when converted back to its original scale (DiGangi et al. 2009).

Hypothesis 3. Population differences in aging rates

For within-group differences (i.e. male-female differences), mean ages at transitions between stages were examined. However, according to Konigsberg et al. (2008), differences in mean age at transition do not necessarily translate into a method developed on one population will not perform well for another population. What directly address this question is evaluating any differences in final age estimates using the probit parameters. For this reason, evaluation of population differences in aging rates was conducted in three steps, using the sex- and population-specific probit parameters obtained in Hypothesis 2. First, a posterior distribution (or age estimates) of each of the traits at a univariate scale was obtained based on a uniform prior. Second, given that the main interest of this study is age estimation at a multivariate scale, posterior distributions of a suit of skeletal traits combined were obtained. In addition, most likely combinations of skeletal traits to be observed for a certain age were estimated for Japanese and

Thai males and females. Lastly, cross-validation of sex- and population-specific probit parameters was conducted. That is, ages at death for Japanese and Thais were estimated by using the probit parameters of the opposite group. For example, Japanese male parameters were used to estimate ages for Thai males, and vice versa. The same procedure was followed for Japanese and Thai females. Accuracy of the age estimates was evaluated using RMSE values (Equation 3) that were calculated using point estimates, modes.

Hypothesis 4. Evaluating pooled-sample parameters

In order to assess any improvement in age estimates when pooled samples were used, multivariate probit analysis parameters were obtained from pooled Asian males (n=295) and females (n=124) that serve as reference samples (Table 6 and **Error! Reference source not found.**). Next, following Konigsberg's (2015) protocol, performance of the two new pooled-sample probit models was evaluated based on simulated new target samples (n=200). To create two new target samples for males and females, 200 ages at death were randomly drawn from a Gompertz mortality function with hypothetical shape parameters of $a_3 = 0.001$ and $b_3 = 0.1$. Given these simulated ages at death, a probable suite of skeletal trait stages was simulated using “rmvnorm” from the package “mvtnorm” (Genz et al. 2014). As a result, two sets of 200 individuals with simulated ages at death along with associated scores for a suite of skeletal traits were generated. These simulated samples enable the new pooled-sample models to be tested on a population that is completely different from the reference Asian sample.

The random walk Metropolis-Hastings sampler of the MCMC methods was used to obtain age-at-death estimates given the simulated skeletal traits of the new target samples. In this step, the pooled-Asian probit parameters and an informed prior (a Gompertz hazard function

with $a_3 = 0.001$ and $b_3 = 0.1$) were used for age estimation of the simulated samples. The simulated ages at death of the target samples served as “documented chronological ages” so that the differences between estimated ages and known ages could be evaluated.

As there was no need for stage-by-stage comparisons for this part of the analysis, the Lagrange multiplier test of goodness-of fit was conducted to evaluate the model fit and the normality assumption. If p-values of the Lagrange multiplier test were significantly below 0.05, suggesting the violation of normality and inadequate fit of the model to the Asian data, the stage collapsing procedure was proceeded using the R functions written by Konigsberg (2015a) and Konigsberg et al. (2016). In addition, pseudo- R^2 for each of the 33 traits was calculated to examine strength of relationship between TA traits and documented ages.

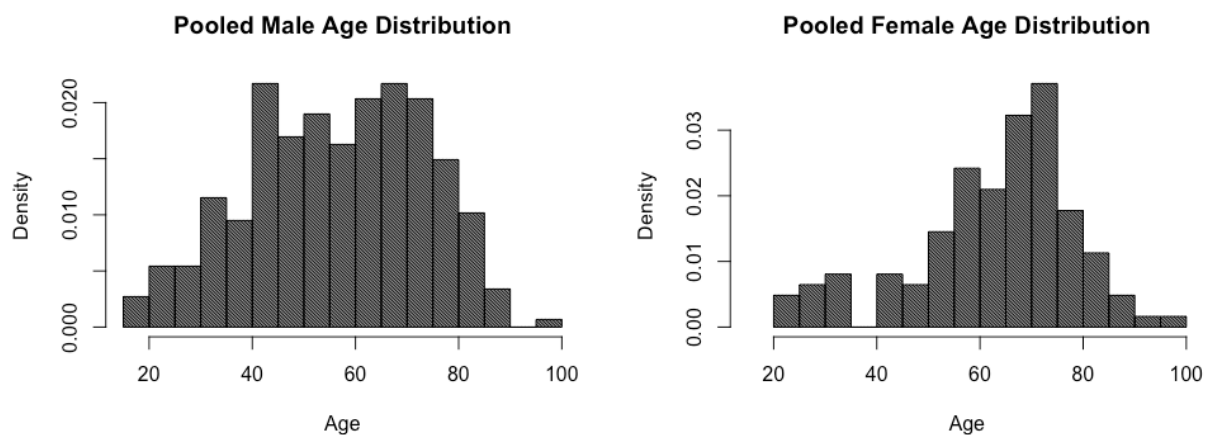


Figure 2. Sample distributions of pooled Asian males and females

Additional Test 1. Accuracy and reliability of conventional methods

In order to evaluate the performance of the conventional aging methods for Asian populations in comparison to TA, the raw scores from the conventional methods were processed the exact same way as the raw scores of TA, except subheading (a) about stage adjustment. That is, the

multivariate ordered probit regression model was fitted to the conventional scores and chronological ages of the Asian samples ($n=259$) to obtain slopes and intercepts. Next, to test how these parameters of the conventional method scores perform, a new target sample was created using the same hypothetical shape parameters of the Gompertz hazard model used in Hypothesis 4 ($a_3= 0.001$ and $b_3= 0.1$). The probit parameters were combined into an informed prior (a Gompertz hazard function with $a_3= 0.001$ and $b_3= 0.1$) to obtain final age estimates on a simulated sample of 200 individuals.

This way of analyzing conventional methods' scores generates age estimates that can be compared well with the age estimates of TA scores than simply using final age estimates from standard descriptive statistics provided with the conventional methods. If standard tables of the conventional methods are used, there will be only six different means of final age estimates, for example, for Suchey-Brooks' method as the method only provides age estimates for its six phases.

Additional Test 2. Retrieving age estimates by combining TA and conventional methods

The best practice guidelines of adult age estimation recommend using more than one age estimation method and examining more than one age indicator to draw final age estimates for each casework (SWGANTH 2013). However, no scientific and convenient way of combining age estimates derived from multiple methods has been proposed so far. In order to test if the analytical framework of the current study can help alleviate this problem, scores of TA and conventional methods were combined and analyzed using the multivariable ordered probit regression combined with an informed prior using the MCMC sampler. A simulated sample was

again used as a new target sample to test the performance of the parameters combining TA and conventional methods.

Chapter 5

Results

This chapter presents results of analyses that were conducted to evaluate the four proposed hypotheses using the four different Japanese and Thai skeletal collections. In the first section, results of an intra-observer error test are presented, followed by tests of the utility of the current TA software, ADBOU, for estimating ages at death of Asian skeletons (Hypothesis 1). The chapter continues to Hypotheses 2-4 that are heavily based on the probit regression model and the MCMC sampler methods. Hypothesis 2 evaluates sex differences in aging by examining differences in mean ages at transition between Japanese males vs. females, and Thai males vs. females. Hypothesis 3 seeks to evaluate the existence of population-specific aging rates and patterns by cross-validating population-specific probit parameters from, for example, Japanese to Thais, and vice versa. Hypothesis 3 builds an argument for weak population-specific aging by presenting a simple univariate ordered probit regression, followed by a quadrivariate ordered probit regression, and a multivariate ordered probit regression. Hypothesis 4 evaluates the impact of pooling two groups to fit a multivariate ordered probit regression model to improve age estimates for Asian populations. The last two additional tests evaluate the utility of three conventional methods (Lovejoy et al. 1985a; Meindl and Lovejoy 1985; Suchey and Katz 1998), and present promising results of using the multivariate ordered probit regression to attain final ages at death from TA and conventional methods combined.

Intra-observer error test

Prior to data analysis, consistency of the observer in scoring age-related skeletal traits was evaluated using the Cohen's weighted kappa. As justified in the previous chapter, intra-observer error was tested only on TA scores. Kappa statistics indicated that most of the TA traits were scored consistently for both observations. Table 7 shows that a majority of trait scores were in moderate to almost perfect agreement. However, lesser agreement was observed for two traits of the right sacroiliac joint, namely superior morphology and apical morphology, with Kappa statistics of 0.406 and 0.407 respectively. Nevertheless, these values still fall within the borderline of fair to moderate agreement categories. Among the three age indicators, scores of the cranial sutures showed the highest agreement to each other. The pubic symphyseal traits were also in comparably high agreement.

Traditionally, it has been reported that the greater the number of stages, the less the repeatability, and therefore reducing the number of stages help increase replicability of the method (Dhanjal et al. 2006; Harris 2007; Johnson 1996b; Olze et al. 2004; Shirley and Ramirez Montes 2015). However, this argument was not applicable for the current case. The ventral margin has the greatest number of stages (7 stages) among all other TA traits, but the trait scores were in almost perfect agreement. Interestingly, despite historical assumptions of poor performance, the scores of cranial sutures with five stages were in the strongest agreement, indicating high repeatability.

Despite that the majority of TA traits were scored consistently by the observer, two traits of the right sacroiliac joint, the superior and apical morphology, showed less agreement. It is possibly due to the subtle nature of sacroiliac joint's characteristics that can be difficult to distinguish, and, therefore, causing differential stage diagnoses at different observations.

Specifically, the sacroiliac joint morphology has five stages to choose from, and the first three stages concern the amount of billows present on the auricular surface: Stage 1 $>2/3$ covered by billows, Stage 2 $1/3$ - $2/3$ covered by billows, and Stage 3 $<1/3$ covered by billows. Although Milner and Boldsen (2013) have quantified the amount of billows better than traditional methods, these three stages still involve approximation that can possibly cause inconsistent scoring. And yet, the inferior morphology of the same side sacroiliac joint had substantial agreement, and the opposite side (left) sacroiliac joint showed moderate to substantial agreement in the superior and apical morphology. This further opens a possibility that, rather than observer error, highly variable trait expressions in superior and apical morphology of the right auricular surface may have complicated correct trait interpretation.

Based on the overall consistency in TA trait scoring of the observer, none of the traits was excluded, and all 33 traits of TA were retained for further analysis.

Hypothesis (1) Testing accuracy and reliability of ADBOU Age Estimation 2.1

It was hypothesized *that the current Transition Analysis method will yield age estimates high in error and bias for Japanese and Thai skeletal samples because the method was originally developed on non-Asian populations*. To evaluate the first hypothesis, the accuracy and reliability of the current TA method was assessed using error and bias as a means of measuring accuracy. The TA program, ADBOU, was used to obtain ages at death of the Asian samples. Maximum likelihood estimates were used as point ages to calculate root mean square error (RMSE, Equation 3 in Chapter 4) and bias (a difference between estimated age and chronological age).

Sex- and population-specific RMSE values of ADBOU suggested that age estimates

Table 7. Weighted Cohen's Kappa statistics and associated p-values

Pubic Symphysis	Left		Right	
Trait	Kappa	p-value	Kappa	p-value
Relief	0.763	1.88E-12	0.799	0
Texture	0.532	2.05E-07	0.592	4.72E-09
Superior Apex	0.662	2.49E-09	0.754	3.99E-12
Ventral Margin	0.837	6.64E-14	0.859	5.77E-15
Dorsal Margin	0.627	1.14E-08	0.693	2.43E-10
Sacroiliac Joint	Left		Right	
Trait	Kappa	p-value	Kappa	p-value
Superior Topography	0.623	1.02E-08	0.572	8.25E-08
Inferior Topography	0.433	2.33E-05	0.464	5.00E-06
Superior Morphology	0.517	8.02E-07	0.406	0.000208
Apical Morphology	0.625	1.26E-08	0.407	0.000177
Inferior Morphology	0.594	7.26E-08	0.74	1.85E-11
Inferior Texture	0.743	1.41E-11	0.86	9.77E-15
Posterior Superior Exostoses	0.702	1.67E-10	0.436	3.77E-08
Posterior Inferior Exostoses	0.724	5.45E-11	0.64	2.34E-12
Posterior Spicules	0.422	7.49E-08	0.46	5.00E-07
Cranial Suture	Left			
Trait	Kappa	p-value		
Coronal Pterica	0.828	5.17E-14		
Sagittal Obelica	0.898	4.44E-16		
Lambdoidal Asterica	0.685	1.80E-11		
Interpalatine	0.844	2.02E-14		
Zygomatico-maxillary	0.753	1.68E-12		

derived from all three age indicators had the lowest error for both Thai and Japanese males and females (Table 8). As for individual age indicators, the pubic symphysis produced the lowest error, followed by the sacroiliac joint and the cranial sutures—an argument supported by the developers of TA (Milner and Boldsen 2012). Although it appeared that the informed prior did not improve point estimates, using the Danish prior did help narrow down age ranges for the Asian samples: An average width of age ranges using a uniform prior was 33 years, while that of the Danish prior was 31 years (Table 27).

Age subgroup comparisons of pooled males and females (Tables 9-10, Figures 3-4) showed that, for both sexes, ADBOU generated the most accurate age estimates for young adults, followed by middle-aged adults and elderly individuals. Figures 3-4 show squared error distributions against known ages, illustrating this tendency. As it was the case for sex- and population-specific RMSE, age estimates of all three age indicators combined produced the most accurate age estimates for all three age subgroups for pooled Asian males and females: 9-11 years for young males, 13-4 years for middle-aged males, 19-21 years for elderly males, 8 years for young females, 14 years for middle-aged females, and 26-27 for old females.

Indicator-specific RMSE values (Tables 9-10) indicated that, for young and middle-aged males, the lowest RMSE was evident in age estimates of the pubic symphysis, while the sacroiliac joint was the second lowest. Cranial sutures had the largest error for young and middle-aged males. For elderly males, however, the pubic symphysis and cranial sutures had lower RMSE (20-21 years) than the sacroiliac joint (27 years).

For young and middle-aged females, the sacroiliac joint had the lowest RMSE (8 years for the young, and 19.5 years for the middle-aged), while the pubic symphysis performed the best for elderly females (RMSE of 20 years) over the other two age indicators, which had 33.5

years of RMSE. For elderly females, adding the sacroiliac joint and cranial sutures to pubic symphysis age estimates seemed to exacerbate the inaccuracy: Regardless of a prior choice, the RMSE of all three elements combined were 26 and 27 years for age estimates using uniform and informed priors for elderly females, whereas error based on the pubic symphysis alone was 20 years.

The bias assessment indicated that, for males, young individuals below 40 years tend to be overestimated, while the underestimation trend becomes more prominent above 60 years. However, for middle-aged males, bias showed a more random distribution. Figure 5 clearly illustrates this (a vertical dashed line was added at 40 and 60 years to mark the three age groups). In addition, Figure 6 further confirms this tendency. This type of figure (sometimes refer to as candlestick charts or open-high-low-close (OHLC) chart in the field of finance) is pioneered in the context of age estimation by Milner and Boldsen (2012), and serves as a convenient way of evaluating a distribution of estimated age ranges and point estimates in relation to known ages. Each of the vertical bars represents an estimated age range (a 95% confidence set) per individual, while a filled dot with an “x” mark is a point age estimate (MLE) and a filled dot stands for the documented chronological age in ascending order. In Figure 6, it is evident that not only point estimates but also age ranges of pooled males are shifted downward relative to the documented ages. For both Japanese and Thai females, there was even stronger underestimation across all age groups (Figure 7) than males. Figure 8 thus illustrates that actual ages (filled dots) are located at the upper end of the vertical bars, and that point estimates (filled dots with x marks) are shifted downward from the known ages, demonstrating that ADBOU underestimates Asian female ages.

Table 8. Population- and sex-specific root mean square error (RMSE, in years)

Age Indicator	Thai		Japanese		Pooled		
	Male	Female	Male	Female	Male	Female	Overall
Pubic Symphysis (Uniform)	19.84	25.43	17.12	21.86	18.36	23.71	20.02
Sacroiliac Joint (Uniform)	25.55	27.66	18.88	27.59	22.09	27.56	23.79
Cranial Sutures (Uniform)	26.46	30.02	27.18	28.66	26.81	29.32	27.58
All (Uniform)	19.80	23.13	14.52	20.96	17.02	22.04	18.63
All (Danish)	19.31	22.66	13.98	20.76	16.51	21.70	18.19

Table 9. Pooled-male age subgroup comparisons of RMSE (in years)

Age Indicator	n	Young (<40yrs)	n	Middle (41-60yrs)	n	Old (61yrs<)
Pubic Symphysis (Uniform)	87	13.67	180	16.84	236	20.29
Sacroiliac Joint (Uniform)	89	14.16	181	17.20	246	27.07
Cranial Sutures (Uniform)	72	22.88	168	28.11	235	20.96
All (Uniform)	93	9.39	183	13.16	248	21.25
All (Danish)	93	11.25	183	13.82	248	19.69

Table 10. Pooled-female age subgroup comparisons of RMSE (in years)

Age Indicator	n	Young (<40yrs)	n	Middle (41-60yrs)	n	Old (61yrs<)
Pubic Symphysis (Uniform)	27	16.04	52	24.68	120	20.29
Sacroiliac Joint (Uniform)	34	8.28	54	19.55	121	33.50
Cranial Sutures (Uniform)	26	12.38	54	24.99	121	33.40
All (Uniform)	35	7.88	56	14.01	126	27.06
All (Danish)	35	8.33	56	14.27	126	26.48

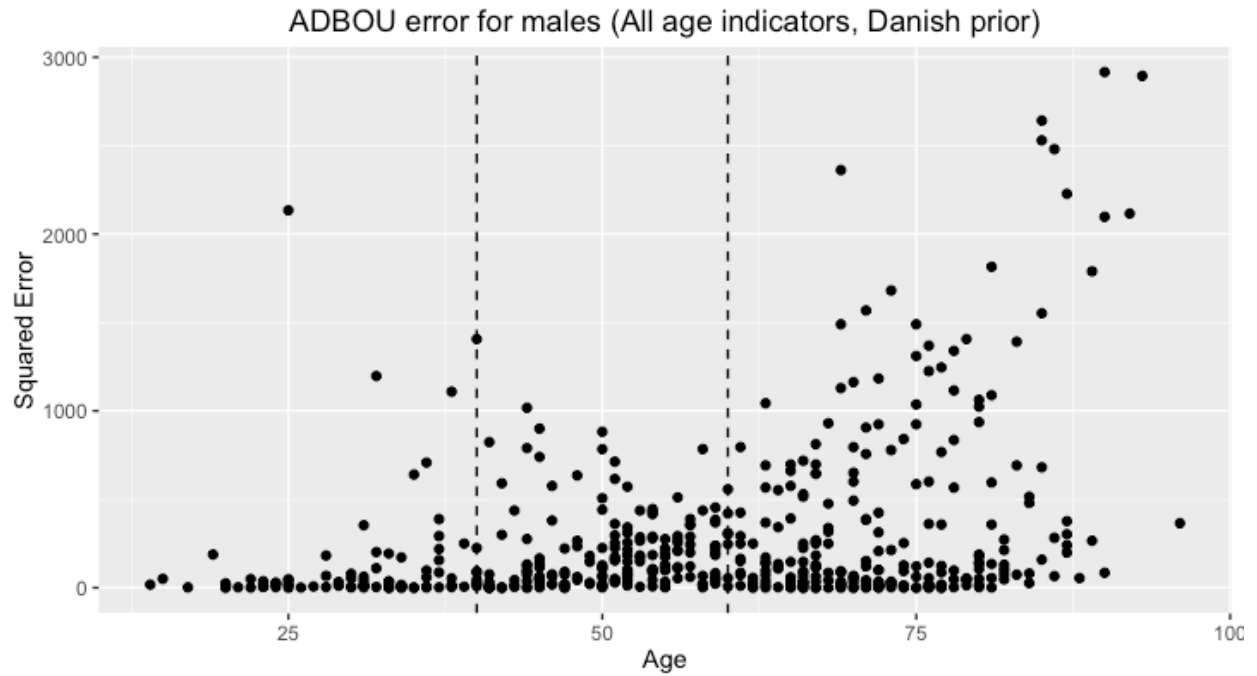


Figure 3. Squared error of ADBOU age estimates for pooled males

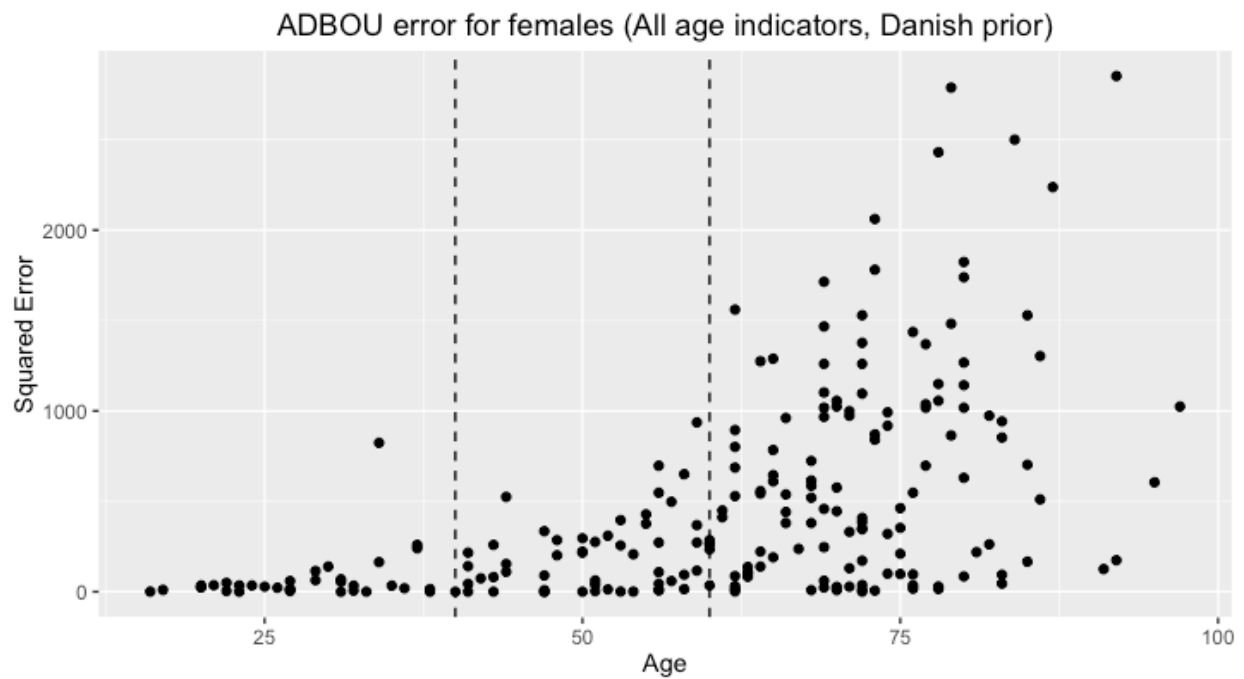


Figure 4. Squared error of ADBOU age estimates for pooled females

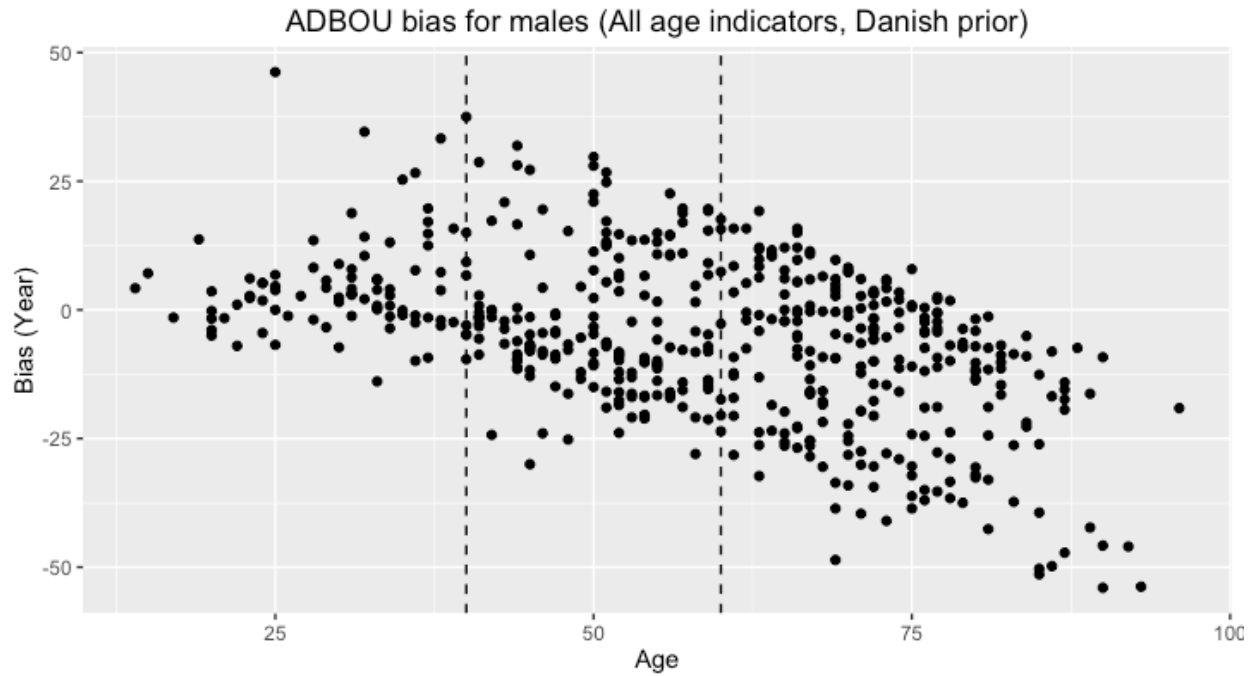


Figure 5. Bias of ADBOU age estimates for pooled males

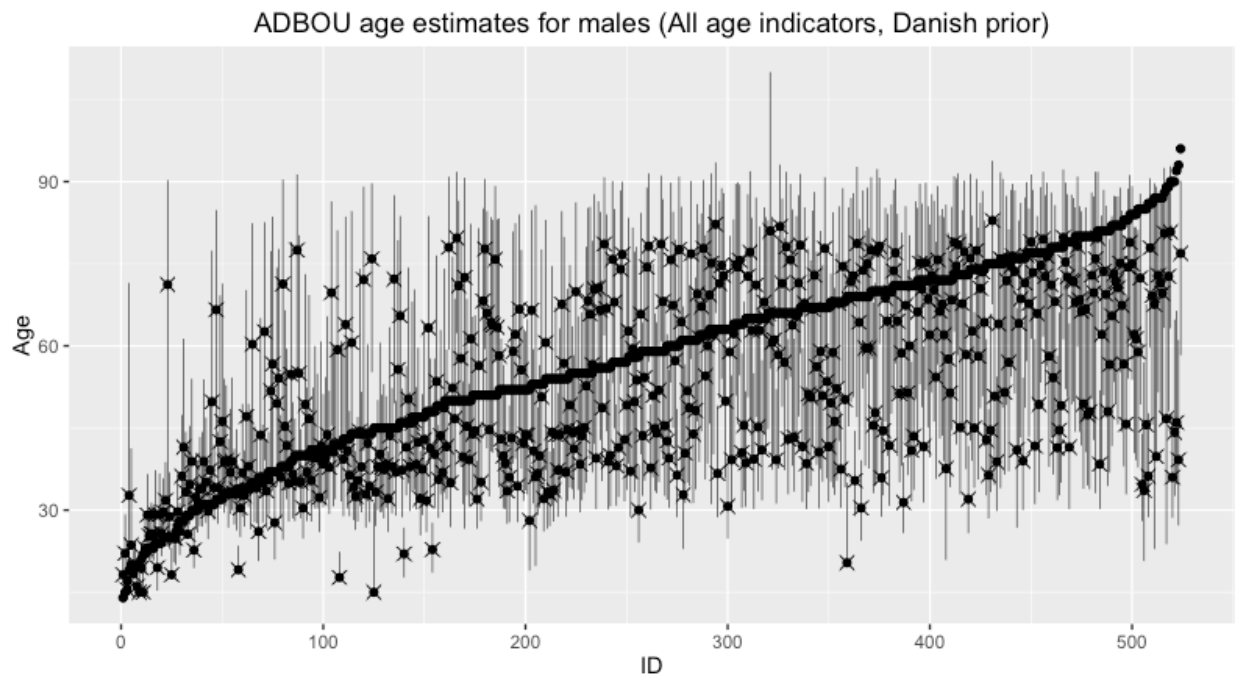


Figure 6. A distribution of age estimates and documented chronological ages for pooled Asian males

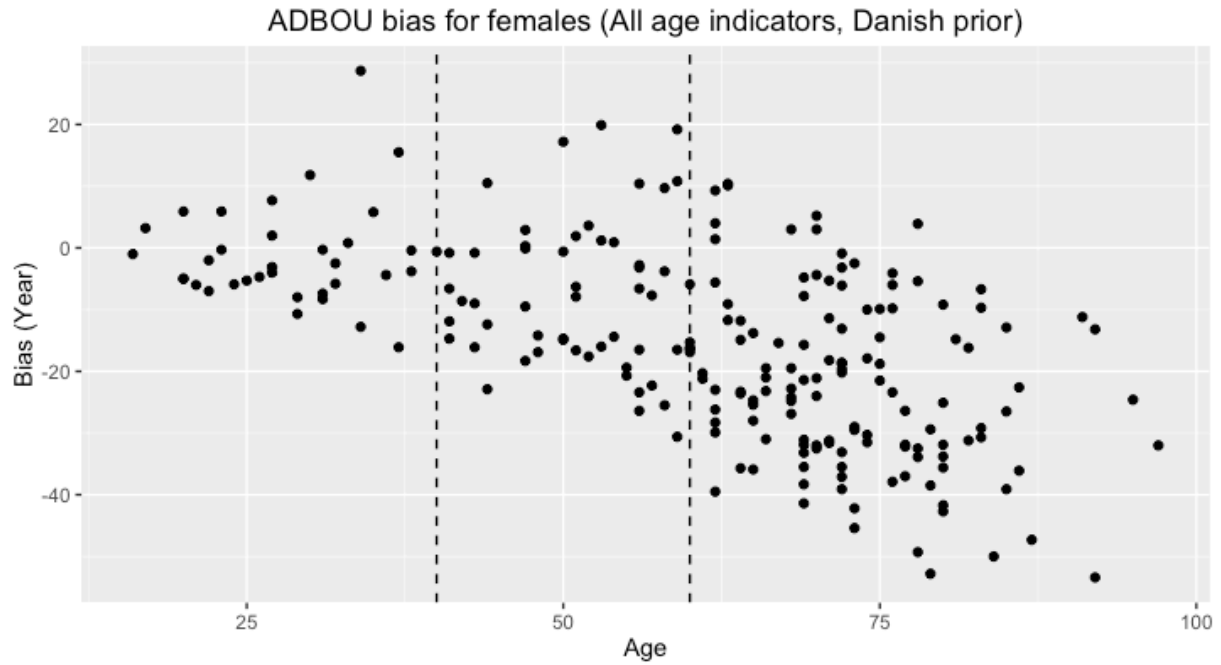


Figure 7. Bias of ADBOU age estimates for pooled females

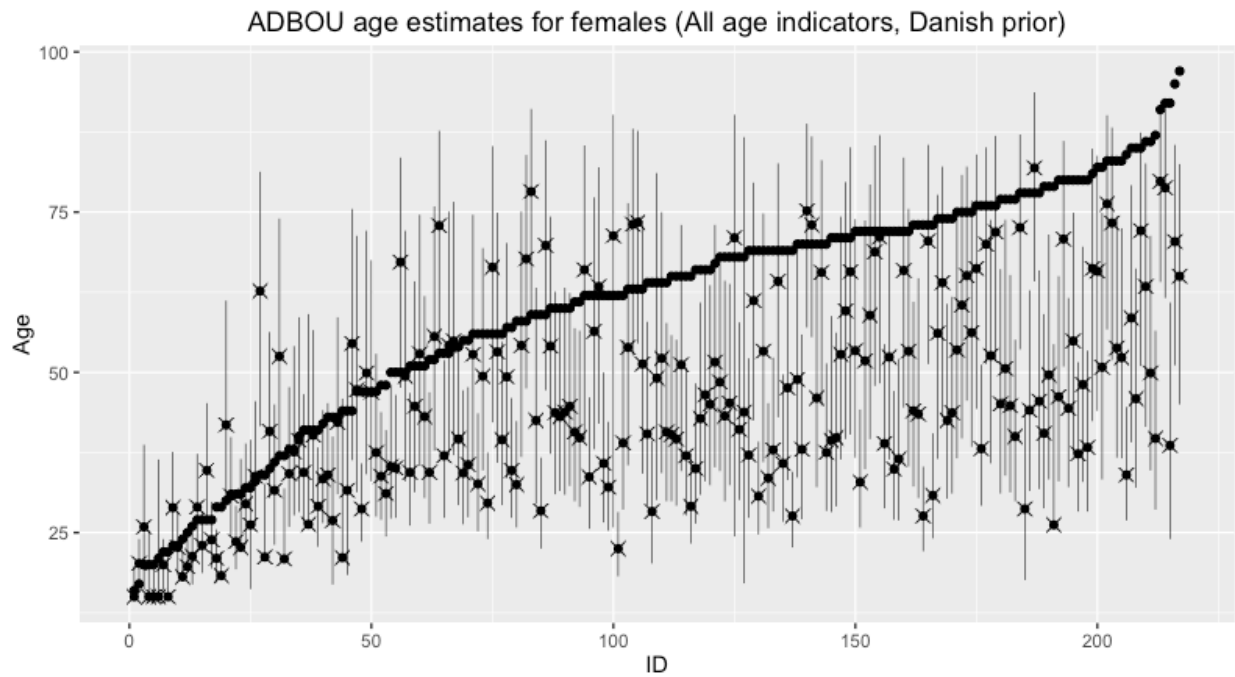


Figure 8. A distribution of age estimates and documented chronological ages for pooled Asian females

Hypothesis (2) Evaluating sex differences in aging rates

The second hypothesis is that *there will be a certain degree of sex differences in age progressive changes between Japanese (or Thai) male skeletons and Japanese (or Thai) female skeletons.*

Differences in aging rates between Japanese males and females, and Thai males and females were evaluated based on mean ages at transition recovered from the probit regression parameters. At the same time, log-normality assumption check was conducted. Log-mean ages at death values as well as mean ages at death in the original scale were examined, and transition distributions are checked for any violation of the normality.

As explained in the methods section, mean ages at transition, as well as standard deviation (SD), can be retrieved using slopes and intercepts of the multivariate ordered probit analysis. Tables 11-14 present a slope and intercepts (or thresholds, as presented as T1, T2, T3, etc. in the table header) for each trait. As described in Chapter 4, each threshold marks a point when a trait transitions from one ordinal stage to next on an underlying continuous normal distribution. It is important to note, therefore, the intercepts or thresholds are not the same as mean ages at transition. Mean ages at transition and standard deviation are recovered using the Equations 4 and 5 in Chapter 4 by applying the probit parameters in Tables 11-14.

The probit analysis provided sex- and population-specific slopes and intercepts for each of the 33 traits for Japanese males and Thai females. However, for Japanese females and Thai males, not all 33 trait parameters were presented due to missing stages despite the stage adjustments. The traits without probit parameters, as indicated with 'NA,' include, for Japanese females, left and right ventral margins, left superior morphology, left posterior inferior exostoses and lambdoidal asterica. For Thai males, left symphyseal texture, left and right ventral margins, and coronal pterica.

The log-normality assumption check

Given the probit regression assumes a log-normal distribution, it is important to check if there is a violation of the assumption before accepting probit regression parameters and mean ages at transition. As justified in Chapter 4, recommendations of Shackelford et al. (2012) were adopted to test Hypotheses 2 and 3. According to Shackelford et al. (2012), when mean ages of transition from one stage to the next are widely distributed, the assumption of log-normality is no longer reasonable. A wide separation of mean ages of transition will lead the cumulative probit to a platykurtotic (flat topped) curve with a large standard deviation. This issue is illustrated in Figure 9 showing stage-by-stage transition distributions of the apical morphology of the left sacroiliac joint for Japanese males and females. As indicated by the peak of the first two distributions on the far left of Figure 9, a log mean transition from Stage 1 to Stage 2 occurs at an early age for both Japanese females (solid line) and males (dashed line). However, the following transition distributions to next higher stages are extremely platykurtotic and almost as flat as a straight line. These distributions hardly conform to normality, and if parameters from these distributions are used to recover mean ages at transitions, it will yield very unrealistic age values, such as 1,926 years of transition from Stage 3 to Stage 4 for Japanese females.

After this inspection step, a number of traits were excluded from the sex- and population-specific comparisons as they yielded unrealistic ages at transition in the original scale, and violated the normality assumption. As a result, only were three traits of the pubic symphysis (i.e. the superior apex, the ventral margin, and the dorsal margin) of Japanese and Thai males and females were used to assess any sex and population difference in aging rates.

Table 15 lists sex- and population-specific log-mean, mode, median and mean ages of transition. Note that the mode, median and mean values are not the same due to the asymmetric

Table 11. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Japanese males (n=138)

Traits	Slope	T1	T2	T3	T4	T5	T6
Symphyseal Relief L	1.349	2.632	3.464	4.435	6.956	NA	NA
Symphyseal Relief R	1.636	3.849	4.572	5.509	8.245	NA	NA
Symphyseal Texture L	1.935	5.718	8.046	NA	NA	NA	NA
Symphyseal Texture R	1.99	6.093	8.242	9.921	NA	NA	NA
Superior Apex L	2.175	6.334	7.783	NA	NA	NA	NA
Superior Apex R	1.735	4.757	5.818	NA	NA	NA	NA
Ventral Margin L	2.854	7.601	8.686	9.097	9.646	10.45	11.908
Ventral Margin R	3.016	8.267	8.958	9.732	10.13	11.032	12.542
Dorsal Margin L	2.43	6.72	7.312	8.574	10.179	NA	NA
Dorsal Margin R	2.516	7.074	7.362	8.805	10.558	NA	NA
Superior Topography L	1.989	6.733	8.074	NA	NA	NA	NA
Superior Topography R	1.2	3.788	4.92	NA	NA	NA	NA
Inferior Topography L	1.992	6.806	8.11	NA	NA	NA	NA
Inferior Topography R	1.359	4.459	5.564	NA	NA	NA	NA
Superior Morphology L	1.223	2.455	4.009	6.257	NA	NA	NA
Superior Morphology R	0.395	-0.673	0.6	3.312	NA	NA	NA
Apical Morphology L	0.45	0.324	1.814	3.188	NA	NA	NA
Apical Morphology R	0.393	-0.271	1.574	2.929	NA	NA	NA
Inferior Morphology L	0.81	1.498	3.297	4.8	NA	NA	NA
Inferior Morphology R	1.036	2.57	3.88	6.087	NA	NA	NA
Inferior Texture L	1.544	6.238	7.493	NA	NA	NA	NA
Inferior Texture R	1.397	5.398	6.652	NA	NA	NA	NA
Posterior Superior Exostoses L	1.053	2.944	5.169	5.3	NA	NA	NA
Posterior Superior Exostoses R	0.965	2.473	4.902	5.098	NA	NA	NA
Posterior Inferior Exostoses L	-0.183	-0.677	1.233	1.932	NA	NA	NA
Posterior Inferior Exostoses R	0.041	0.244	1.791	2.39	NA	NA	NA
Posterior Spicules L	0.506	2.874	NA	NA	NA	NA	NA
Posterior Spicules R	0.151	1.358	3.08	NA	NA	NA	NA
Coronal Pterica	1.149	2.844	3.695	4.634	5.15	NA	NA
Sagittal Obelica	1.142	3.202	3.538	4.731	5.609	NA	NA
Lambdoidal Asterica	0.655	1.753	2.668	5.065	NA	NA	NA
Interpalatine	1.79	5.772	6.626	7.711	NA	NA	NA
Zygomaticomaxillary	0.894	2.113	3.057	4.284	5.62	NA	NA

Table 12. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Japanese females (n=45)

Trait	Slope	T1	T2	T3	T4	T5	T6
Symphyseal Relief L	1.168	2.538	3.327	4.206	6.207	NA	NA
Symphyseal Relief R	1.479	3.632	4.252	5.25	7.425	NA	NA
Symphyseal Texture L	1.331	4.087	5.17	6.583	NA	NA	NA
Symphyseal Texture R	0.764	1.74	3.018	4.476	NA	NA	NA
Superior Apex L	1.741	4.564	6.163	NA	NA	NA	NA
Superior Apex R	1.421	3.653	4.889	NA	NA	NA	NA
Ventral Margin L*	-1.851	5.128	5.439	NA	6.217	7.378	7.816
Ventral Margin R*	-2.524	7.512	7.868	NA	8.397	9.845	10.734
Dorsal Margin L	1.39	2.971	4.324	4.835	5.913	NA	NA
Dorsal Margin R	1.604	3.498	4.601	5.957	6.779	NA	NA
Superior Topography L	0.894	2.882	3.823	NA	NA	NA	NA
Superior Topography R	1.37	4.577	5.843	NA	NA	NA	NA
Inferior Topography L	0.6	1.97	3.038	NA	NA	NA	NA
Inferior Topography R	0.857	3.108	3.977	NA	NA	NA	NA
Superior Morphology L	NA	NA	NA	NA	NA	NA	NA
Superior Morphology R	NA	NA	NA	NA	NA	NA	NA
Apical Morphology L	0.659	1.886	3.202	4.158	NA	NA	NA
Apical Morphology R	0.863	2.425	3.629	4.537	NA	NA	NA
Inferior Morphology L	1.379	4.254	5.975	7.159	NA	NA	NA
Inferior Morphology R	1.286	4.072	5.368	NA	NA	NA	NA
Inferior Texture L	1.99	8.117	9.165	NA	NA	NA	NA
Inferior Texture R	1.689	6.879	7.709	NA	NA	NA	NA
Posterior Superior Exostoses L	1.148	3.945	5.741	6.699	NA	NA	NA
Posterior Superior Exostoses R	0.92	3.051	4.783	5.683	NA	NA	NA
Posterior Inferior Exostoses L	NA	NA	NA	NA	NA	NA	NA
Posterior Inferior Exostoses R	-0.694	-3.509	-2.186	-0.793	NA	NA	NA
Posterior Spicules L	0.619	3.628	4.427	NA	NA	NA	NA
Posterior Spicules R	1.007	4.945	5.664	NA	NA	NA	NA
Coronal Pterica	1.763	5.585	6.356	7.213	7.504	NA	NA
Sagittal Obelica	1.613	5.482	6.139	7.335	7.783	NA	NA
Lambdoidal Asterica	NA	NA	NA	NA	NA	NA	NA
Interpalatine	1.507	6.234	6.63	7.469	NA	NA	NA
Zygomaticomaxillary	0.946	2.714	4.076	5.396	NA	NA	NA

*Ventral margin skips from Stage 2 to Stage 4 due to absence of Stage 3

Table 13. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Thai males (n=157)

Trait	Slope	T1	T2	T3	T4	T5	T6
Symphyseal Relief L	0.999	1.091	1.865	3.362	5.828	NA	NA
Symphyseal Relief R	0.848	0.11	1.137	2.589	5.713	NA	NA
Symphyseal Texture L	NA	NA	NA	NA	NA	NA	NA
Symphyseal Texture R	0.242	-1.437	1.001	2.805	NA	NA	NA
Superior Apex L	1.108	2.492	3.726	NA	NA	NA	NA
Superior Apex R	1.569	4.314	5.437	NA	NA	NA	NA
Ventral Margin L*	- 2.067	NA	6.013	6.507	6.798	7.407	9.059
Ventral Margin R*	-2.228	NA	6.638	7.029	7.197	7.817	9.816
Dorsal Margin L	2.221	5.244	6.602	7.924	9.787	NA	NA
Dorsal Margin R	1.993	4.611	5.513	6.941	8.751	NA	NA
Superior Topography L	0.657	1.09	2.766	NA	NA	NA	NA
Superior Topography R	-0.102	-2.029	-0.318	NA	NA	NA	NA
Inferior Topography L	0.268	-0.483	1.552	NA	NA	NA	NA
Inferior Topography R	0.022	-1.328	0.445	NA	NA	NA	NA
Superior Morphology L	0.373	-0.812	0.694	3.729	NA	NA	NA
Superior Morphology R	0.468	-0.779	1.149	3.617	NA	NA	NA
Apical Morphology L	0.465	0.455	2.117	3.515	NA	NA	NA
Apical Morphology R	0.643	0.911	2.66	4.224	NA	NA	NA
Inferior Morphology L	1.29	3.598	4.944	7.069	NA	NA	NA
Inferior Morphology R	0.79	1.77	3.06	4.971	NA	NA	NA
Inferior Texture L	1.573	6.492	7.85	NA	NA	NA	NA
Inferior Texture R	2.393	9.904	11.49	NA	NA	NA	NA
Posterior Superior Exostoses L	0.937	2.461	4.664	4.75	NA	NA	NA
Posterior Superior Exostoses R	1.148	3.587	5.504	5.594	NA	NA	NA
Posterior Inferior Exostoses L	0.561	2.847	4.325	4.591	NA	NA	NA
Posterior Inferior Exostoses R	0.84	3.839	5.401	5.778	NA	NA	NA
Posterior Spicules L	1.394	7.185	8.193	NA	NA	NA	NA
Posterior Spicules R	1.537	7.633	8.296	NA	NA	NA	NA
Coronal Pterica	NA	NA	NA	NA	NA	NA	NA
Sagittal Obelica	-0.023	-2.062	-1.656	-0.024	1.046	NA	NA
Lambdoidal Asterica	0.317	-0.334	0.812	3.237	3.884	NA	NA
Interpalatine	1.025	2.21	3.683	4.902	NA	NA	NA
Zygomaticomaxillary	0.299	-0.84	0.136	1.744	3.445	NA	NA

* No Stage 1 for the ventral margin

Table 14. Slopes and intercepts (Threshold 1-6 [T1-6]) for the multivariate ordered probit model: Thai females (n=79)

Trait	Slope	T1	T2	T3	T4	T5	T6
Symphyseal Relief L	1.294	2.087	2.995	5.017	6.777	NA	NA
Symphyseal Relief R	1.462	2.871	4.208	5.673	7.716	NA	NA
Symphyseal Texture L	1.926	5.638	7.68	9.769	NA	NA	NA
Symphyseal Texture R	1.494	4.042	5.773	7.809	NA	NA	NA
Superior Apex L	3.307	11.131	12.845	NA	NA	NA	NA
Superior Apex R	3.499	11.563	13.419	NA	NA	NA	NA
Ventral Margin L	3.038	9.257	10.184	11.565	11.683	12.403	12.954
Ventral Margin R	2.362	4.931	7.844	8.436	8.612	9.437	9.865
Dorsal Margin L	2.005	5.47	6.933	7.872	8.995	NA	NA
Dorsal Margin R	1.881	4.256	6.115	7.129	8.292	NA	NA
Superior Topography L	1.528	5.009	6.436	NA	NA	NA	NA
Superior Topography R	1.131	3.531	4.846	NA	NA	NA	NA
Inferior Topography L	2.31	8.47	9.742	NA	NA	NA	NA
Inferior Topography R	1.738	6.348	7.645	NA	NA	NA	NA
Superior Morphology L	0.852	1.825	2.659	5.991	NA	NA	NA
Superior Morphology R	NA	NA	NA	NA	NA	NA	NA
Apical Morphology L	0.657	1.62	3.065	4.371	NA	NA	NA
Apical Morphology R	0.927	2.853	4.012	5.752	NA	NA	NA
Inferior Morphology L	1.493	4.766	6.173	8.378	NA	NA	NA
Inferior Morphology R	1.632	5.154	6.949	8.655	NA	NA	NA
Inferior Texture L	2.03	8.372	9.202	NA	NA	NA	NA
Inferior Texture R	1.574	6.598	7.692	NA	NA	NA	NA
Posterior Superior Exostoses L	0.715	2.195	3.812	4.022	NA	NA	NA
Posterior Superior Exostoses R	1.486	5.405	7.076	7.364	NA	NA	NA
Posterior Inferior Exostoses L	0.508	1.665	2.558	3.1	NA	NA	NA
Posterior Inferior Exostoses R	0.061	-0.413	0.618	1.303	NA	NA	NA
Posterior Spicules L	0.464	2.934	3.987	NA	NA	NA	NA
Posterior Spicules R	0.425	2.502	3.494	NA	NA	NA	NA
Coronal Pterica	0.898	1.894	2.592	3.716	4.186	NA	NA
Sagittal Obelica	-0.077	-1.397	-0.915	0.558	1.151	NA	NA
Lambdoidal Asterica	1.011	3.381	4.154	5.931	6.504	NA	NA
Interpalatine	2.046	8.036	8.736	9.547	NA	NA	NA
Zygomaticomaxillary	0.181	-0.316	0.9	2.911	NA	NA	NA

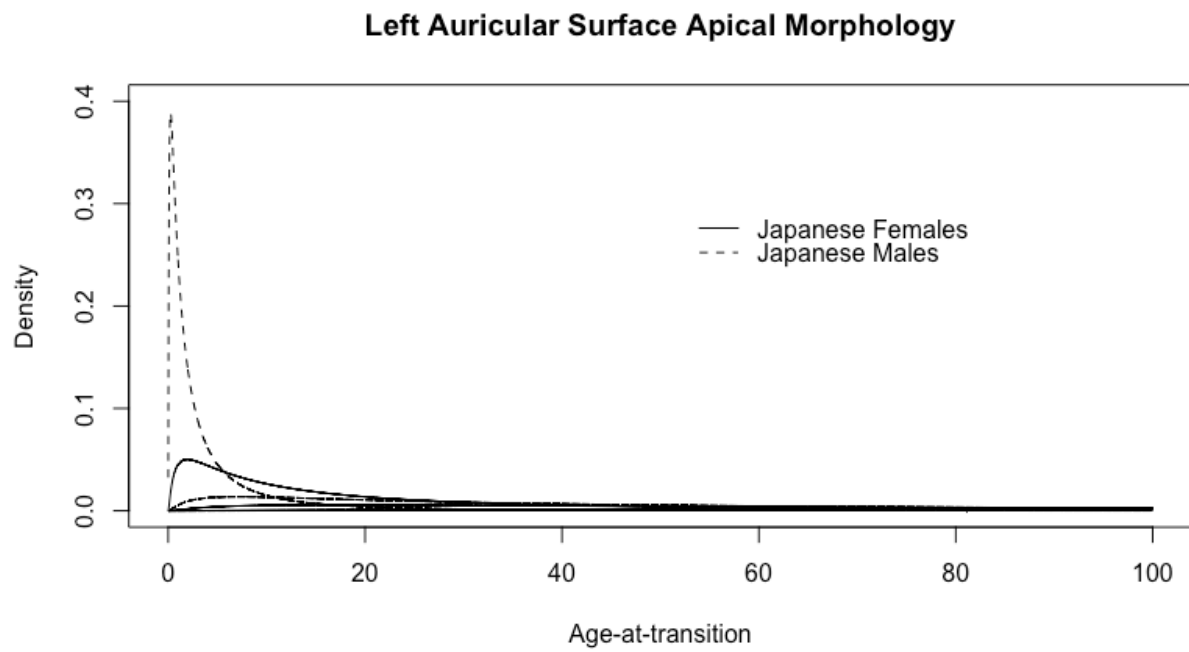


Figure 9. Stage-by-stage transition distributions of the apical morphology of the left sacroiliac joint for Japanese males and females

nature of the log-normal distribution (DiGangi et al. 2009). Although the ventral margin of the Japanese females and Thai males had a few absent stages, probit regression parameter values and mean ages at transition with SDs were still presented in Table 15 because this trait was one of the few traits that met the assumptions. In addition to the three pubic symphyseal traits, Japanese males had extra traits that could be added to the sex- and population-specific analysis for Hypothesis 3. Those traits include the superior and inferior demiface topography of the left and right sacroiliac joints, and interpalatine sutures. For the same reason, the left superior and inferior demiface topography were retained for Thai females for testing Hypothesis 3.

Japanese males vs. females

For Japanese male-female comparisons, the schedule of transition between stages showed a mixed pattern of similarity and differences depending on traits and stages within a trait. For example, mean ages at transition of the superior apex of the left pubic symphysis are almost identical for both Japanese males and females: A transition from Stage 1 to 2, and from Stage 2 to 3 occurs at 19 and 37 years respectively for Japanese females, and 21.5 and 37 years for Japanese males (Table 15 and Figure 10). Figure 10 shows significant overlaps of these four transition distributions. However, the right side superior apex showed different mean ages at transition for Japanese males and females that seem to be as a result of asymmetric scoring in Japanese males (Table 15).

Generally speaking, for those traits showing sex differences, Japanese females tend to experience either a faster or similar timing of transition for the first stage, but, from the second stage onward, females have a constantly delay in transition. For example, a transition from Stage 1 to 2 of the left ventral margin occurs at 18.5 and 19 years for Japanese females and males

respectively. Yet, skipping Stage 2-4 due to absence of Stage 3 for Japanese females, the magnitude of a difference between males and females becomes large from the transition to Stage 5: Specifically, Japanese females show much slower transition than males.

The greatest difference was observed at a transition from Stage 5 to 6 (from “Rampart completion” to “Rim”). For Japanese females, the transition occurs at 62 years, while Japanese males undergo at a much younger age of 37 years, resulting a 25-year difference in transition. The dorsal margins showed the same trend: Japanese females experienced a faster first transition, followed by constant delay in later transitions than males. It appears that the amount of a delay in stage transition of Japanese females becomes large toward higher stages.

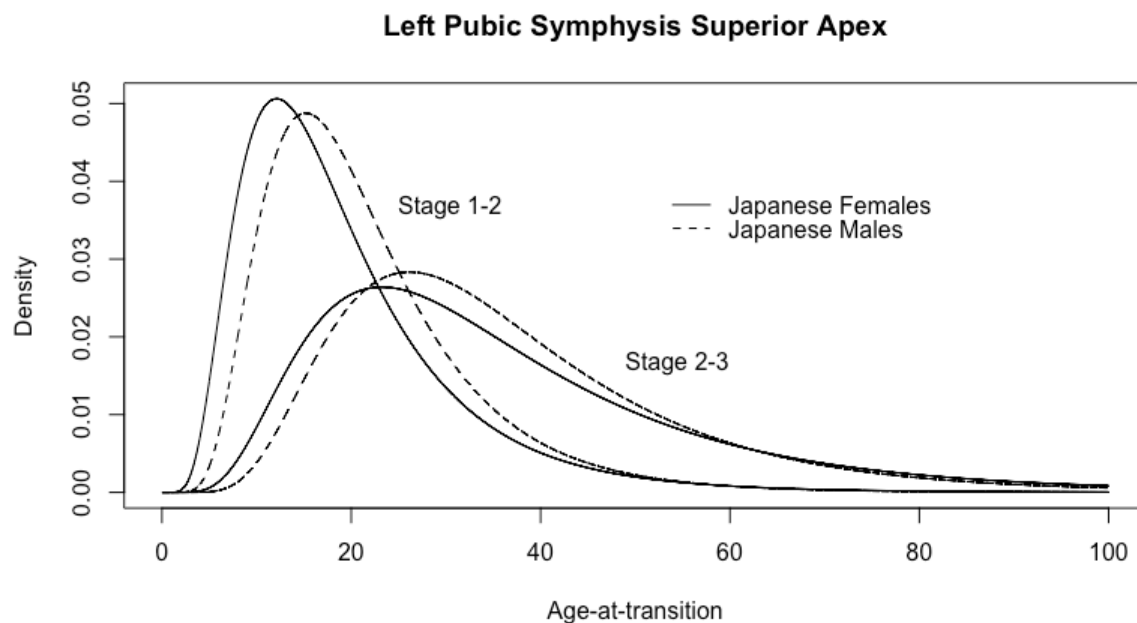


Figure 10. Transition distributions for the left superior apex of Japanese males and females

Thai males vs. females

A mixed pattern of mean ages at transition for Thais was also observed. However, differences in ages at transition were greater within Thais than Japanese (Table 15). For example, ages at

transition from Stage 1-2, and Stage 2-3 for the superior apex of the left pubic symphysis is 30 and 48 years for Thai females, while it is 16.5 and 41 years for Thai males. Unlike Japanese females, the transition from Stage 1 to 2 of the left superior apex for Thai females was notably later, with a 13.5-year difference, than Thai males. For the ventral margins, Thai females experience ages at transition at older ages than Thai males (Table 15), with an exception of the last transition of the ventral margin to the “Breakdown” stage (74-77 years for Thai females and 80-91 years for Thai males). Comparisons of the first transition for the ventral margin could not be made due to absence of Stage 1 in Thai males. For the dorsal margins, Thai females showed either a similar or faster timing of transition for Stage 1, depending on the side, but significantly delayed mean ages at transition in later stages become apparent (Table 15). Figure 11 illustrates that male-female differences in the dorsal margin are particularly marked at Stage 2-3 and Stage 3-4 transitions.

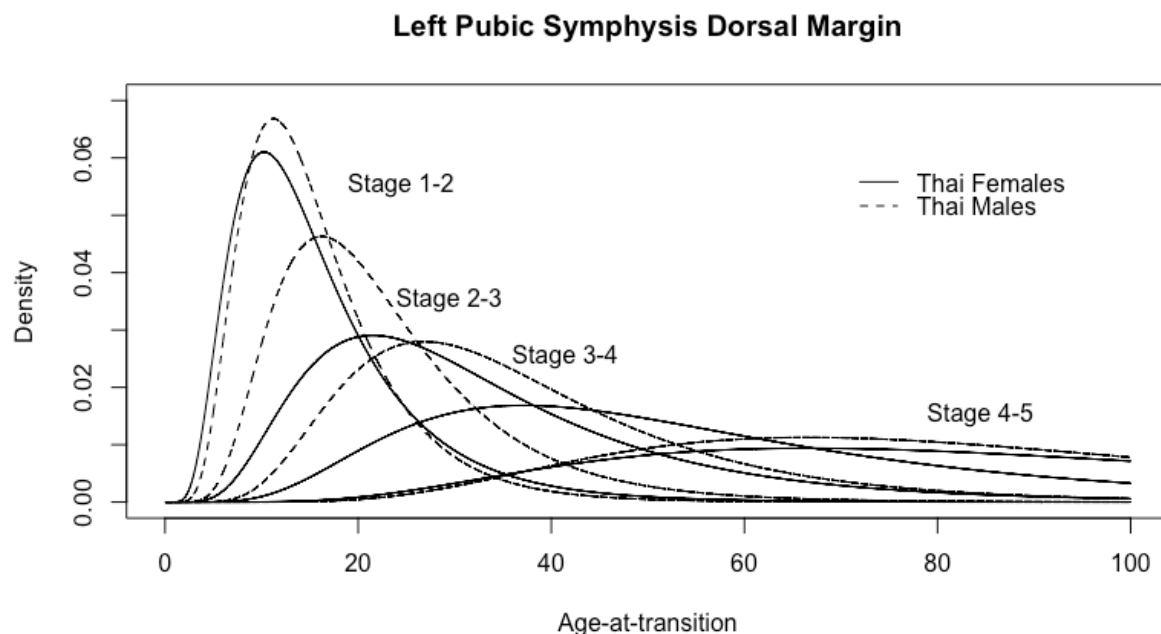


Figure 11. Transition distributions for the left dorsal margin of Thai males and females

Table 15. Mean ages at transition and standard deviations (SD) of three pubic symphyseal traits
(1) Superior Apex

	Left					Right				
Japanese Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	2.805	12.1	16.5	19.3	1-2	2.656	9.1	14.2	17.8
	2-3	3.456	23.2	31.7	37	2-3	3.382	18.9	29.4	36.7
	SD	0.557				SD	0.665			
Japanese Males		lnMu	Modal	Median	Mean		lnMu	Modal	Median	Mean
	1-2	2.951	15.2	19.1	21.5	1-2	2.843	12.8	17.2	19.9
	2-3	3.494	26.2	32.9	36.9	2-3	3.317	20.5	27.6	32
	SD	0.48				SD	0.545			
	Left					Right				
Thai Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	3.354	25.4	28.6	30.4	1-2	3.359	25.5	28.8	30.5
	2-3	3.818	40.4	45.5	48.3	2-3	3.726	36.8	41.5	44.1
	SD	0.346				SD	0.345			
Thai Males		lnMu	Modal	Median	Mean		lnMu	Modal	Median	Mean
	1-2	2.432	5.4	11.4	16.5	1-2	2.899	12.8	18.2	21.6
	2-3	3.347	13.6	28.4	41.1	2-3	3.47	22.7	32.1	38.2
	SD	0.86				SD	0.589			

Table 15 (Continued)
(2) The ventral margin

	Left					Right				
Japanese Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	2.771	11.9	16	18.5	1-2	2.976	16.8	19.6	21.2
	2-4*	2.938	14.1	18.9	21.8	2-4	3.117	19.3	22.6	24.4
	4-5	3.358	21.5	28.7	33.2	4-5	3.326	23.8	27.8	30.1
	5-6	3.986	40.2	53.8	62.3	5-6	3.9	42.2	49.4	53.4
	6-7	4.222	50.9	68.2	78.9	6-7	4.253	60.1	70.3	76.1
	SD	0.54				SD	0.396			
Japanese Males		lnMu	Modal	Median	Mean		lnMu	Modal	Median	Mean
	1-2	2.883	15.4	17.9	19.2	1-2	2.923	16.4	18.6	19.8
	2-3	3.081	18.8	21.8	23.5	2-3	3.045	18.5	21	22.4
	3-4	3.14	19.9	23.1	24.9	3-4	3.191	21.4	24.3	25.9
	4-5	3.256	22.4	25.9	27.9	4-5	3.261	22.9	26.1	27.8
	5-6	3.541	29.7	34.5	37.2	5-6	3.554	30.7	35	37.3
	6-7	4.227	59	68.5	73.8	6-7	4.201	58.7	66.8	71.2
	SD	0.386				SD	0.358			
	Left					Right				
Thai Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	3.059	17.7	21.3	23.4	1-2	2.539	9	12.7	15
	2-3	3.243	21.3	25.6	28.1	2-3	3.113	16	22.5	26.7
	3-4	3.68	32.9	39.6	43.5	3-4	3.342	20.1	28.3	33.6
	4-5	3.706	33.8	40.7	44.7	4-5	3.395	21.2	29.8	35.4
	5-6	4.022	46.3	55.8	61.2	5-6	3.875	34.2	48.2	57.2
	6-7	4.261	58.9	70.9	77.8	6-7	4.133	44.3	62.4	74
	SD	0.431				SD	0.585			
Thai Males	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2*	NA	NA	NA	NA	1-2*	NA	NA	NA	NA
	2-3	2.908	14.5	18.3	20.6	2-3	2.979	16.1	19.7	21.8
	3-4	3.148	18.4	23.3	26.2	3-4	3.155	19.2	23.5	25.9
	4-5	3.288	21.2	26.8	30.1	4-5	3.23	20.7	25.3	28
	5-6	3.583	28.5	36	40.4	5-6	3.508	27.3	33.4	36.9
	6-7	4.382	63.3	80	89.9	6-7	4.406	67	81.9	90.6
	SD	0.484				SD	0.449			

*Stage 3 is absent in the Japanese female sample, and Stage 1 is absent in the Thai male sample

Table 15 (Continued)
(3) The dorsal margin

	Left					Right				
Japanese Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	2.292	5.8	9.9	12.9	1-2	2.418	7.6	11.2	13.7
	2-3	3.011	12	20.3	26.4	2-3	2.872	11.9	17.7	21.5
	3-4	3.392	17.5	29.7	38.7	3-4	3.624	25.3	37.5	45.7
	4-5	4.267	42.1	71.3	92.9	4-5	4.234	46.5	69	84.1
	SD	0.727				SD	0.629			
Japanese Males		lnMu	Modal	Median	Mean		lnMu	Modal	Median	Mean
	1-2	2.875	14.6	17.7	19.5	1-2	2.895	15.2	18.1	19.7
	2-3	2.996	16.5	20	22	2-3	2.937	15.9	18.9	20.5
	3-4	3.433	25.6	31	34.1	3-4	3.416	25.7	30.4	33.2
	4-5	4.223	56.4	68.2	75.1	4-5	4.233	58.1	68.9	75.1
	SD	0.437				SD	0.414			
	Left					Right				
Thai Females	Stage	lnMu	Modal	Median	Mean	Stage	lnMu	Modal	Median	Mean
	1-2	2.628	10.2	13.8	16.1	1-2	2.403	7.2	11.1	13.7
	2-3	3.37	21.5	29.1	33.8	2-3	3.098	14.4	22.2	27.5
	3-4	3.915	37.1	50.1	58.3	3-4	3.683	25.8	39.8	49.3
	4-5	4.5	66.6	90	104.7	4-5	4.434	54.7	84.3	104.5
	SD	0.549				SD	0.657			
Thai Males		lnMu	Modal	Median	Mean		lnMu	Modal	Median	Mean
	1-2	2.644	11.2	14.1	15.7	1-2	2.566	10	13	14.8
	2-3	3.011	16.2	20.3	22.7	2-3	2.858	13.4	17.4	19.9
	3-4	3.515	26.8	33.6	37.6	3-4	3.433	23.8	31	35.3
	4-5	4.424	66.6	83.4	93.4	4-5	4.402	62.7	81.6	93.1
	SD	0.475				SD	0.513			

A mixed pattern in the schedule of skeletal aging between sexes

Based on the three symphyseal traits, the mean ages at transition for males and females demonstrated a mixed pattern in the transition ages for both Japanese and Thais (Table 15). Females experience the first transition earlier than males or the two sexes shared a similar schedule. After the first transition, there was a delay in transition to next higher stages in females. However, exceptionally, the Japanese males and females showed the same schedule of transitions for the left superior apex of the pubic symphysis.

Another exception was the transition to the last stage of the ventral margin for Thai females: Thai females experienced faster transition of the last stage of the ventral margin degeneration. To explain why such mixed pattern emerges, there are both biological and technical reasons. However, given that there are considerable sample size differences between males and females, sampling error cannot be confidently ruled out. Further discussions on this subject matter are presented in the following chapter.

Hypothesis (3) Population differences in aging rates and its impact on age estimation

The third hypothesis is that *an age estimation model developed on Japanese will not necessarily produce inaccurate age estimates for Thais, and vice versa*. Recall that it was argued that differences in aging rates between two populations do not necessarily lead to inaccurate age estimates when a method developed on one population is used for the other (Konigsberg et al. 2008). The question of whether or not differences in Japanese- and Thai-specific aging rates in Table 15 yield inaccurate final age estimates was conducted in three steps. As this part of the analysis is particularly complicated, the three steps explained in the previous chapter are summarized again. First, a posterior distribution of each of the above symphyseal traits was

obtained based on a uniform prior. This analysis assumes that we are estimating ages of Japanese and Thais using a single trait. Second, given that the main approach of this study is multivariate, posterior distributions (or age estimates) of all symphyseal traits combined were obtained. In this step, we are assuming that we are estimating ages of Japanese and Thais using four symphyseal traits. In addition, most likely combinations of skeletal traits to be observed for a certain age were estimated for Japanese and Thai males and females. For the first and second steps, left and right superior apices and dorsal margins were used, and the ventral margins are excluded due to absence of some stages and incomplete parameters values. Thus, age estimation was made using probit parameters of the same group.

Lastly, cross-validation of sex- and population-specific probit parameters was conducted based on the all four pubic symphyseal traits above, and additional traits retained for Japanese males and Thai females based on the log-normality check in Hypothesis 2. In addition to the three pubic symphyseal traits, Japanese males yielded probit parameters from the superior and inferior demiface topography of the left and right sacroiliac joints, and interpalatine sutures that can be applied to Thai males. For Thai females, the four pubic symphyseal traits and the left superior and inferior demiface topography were added to retrieve probit parameters for cross-validation of the Thai female parameters on Japanese females. More specifically, for Japanese male age estimation using the multivariate ordered probit regression, Thai male parameters of four pubic symphyseal traits, the superior apex and dorsal margin of the left and right pubic symphysis, were used to obtain age estimates for Japanese males. The ventral margin of the pubic symphysis was excluded given the absence of stage 1 in the Thai male sample. Ages at death of Thai males were estimated using Japanese male parameters based on the following 11 traits: superior apex, ventral margin, and dorsal margin, of the left and right pubic symphysis,

and the superior and inferior demiface topography of the left and right sacroiliac joints, and interpalatine sutures. To estimate ages at death of Japanese females, Thai female parameters were derived from the following eight traits: superior apex, ventral margin and dorsal margin of the left and right pubic symphyses, and the superior and inferior demiface topography of the left sacroiliac joints. To obtain final age estimates for Thai females, four traits of Japanese females were used. The traits include the left and right superior apices, and dorsal margins. Again the ventral margin was excluded due to absence of Stage 3 in the Japanese female sample.

Univariate analysis

Figure 12 through 19 directly compare posterior distributions of each of the trait stages for Japanese males vs. Thai males, and Japanese females vs. Thai females. Each distribution represents probabilities of being at certain ages given the trait stage. A dotted horizontal line indicates an age range equivalent to a 95% confidence set in a frequentist term (Konigsberg and Frankenberg 2002). The figures demonstrate that despite the differences we saw in the mean ages at transition in Table 15, there is a little influence of such differences to posterior probabilities as a considerable overlap in age estimates of Japanese and Thai males shows (Figures 12-14). The only exception was the right dorsal margin (Figure 15).

On the other hand, the female groups displayed less overlaps in their age estimates (Figures 16-18). Notice that the Thai female distributions (solid-line) and the Japanese female distributions (dashed-line) are distinctly apart from each other in Figures 16-18. However, the right dorsal margin yielded age estimates that significantly overlap for Japanese females and Thai females (Figure 19).

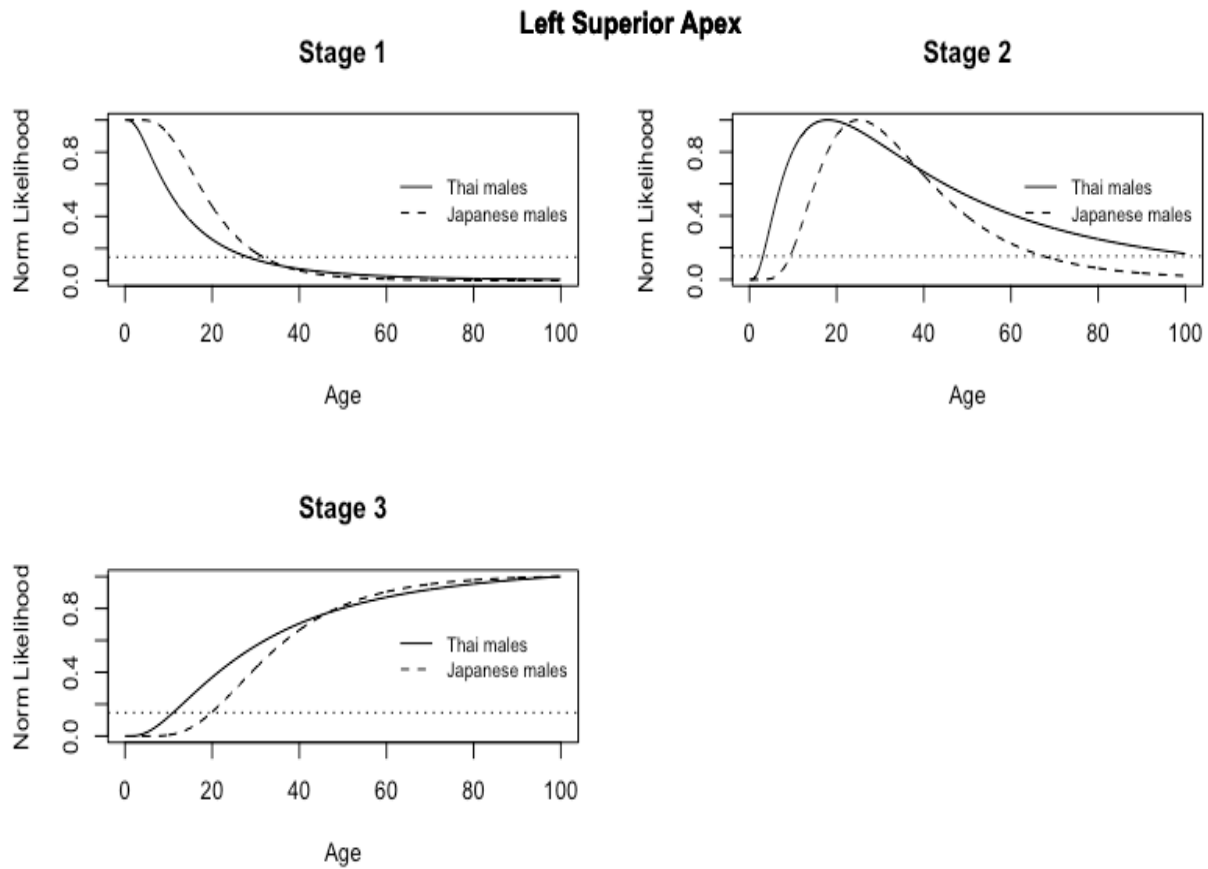


Figure 12. The probabilities of being in each stage of the left superior apex: Japanese and Thai male comparisons

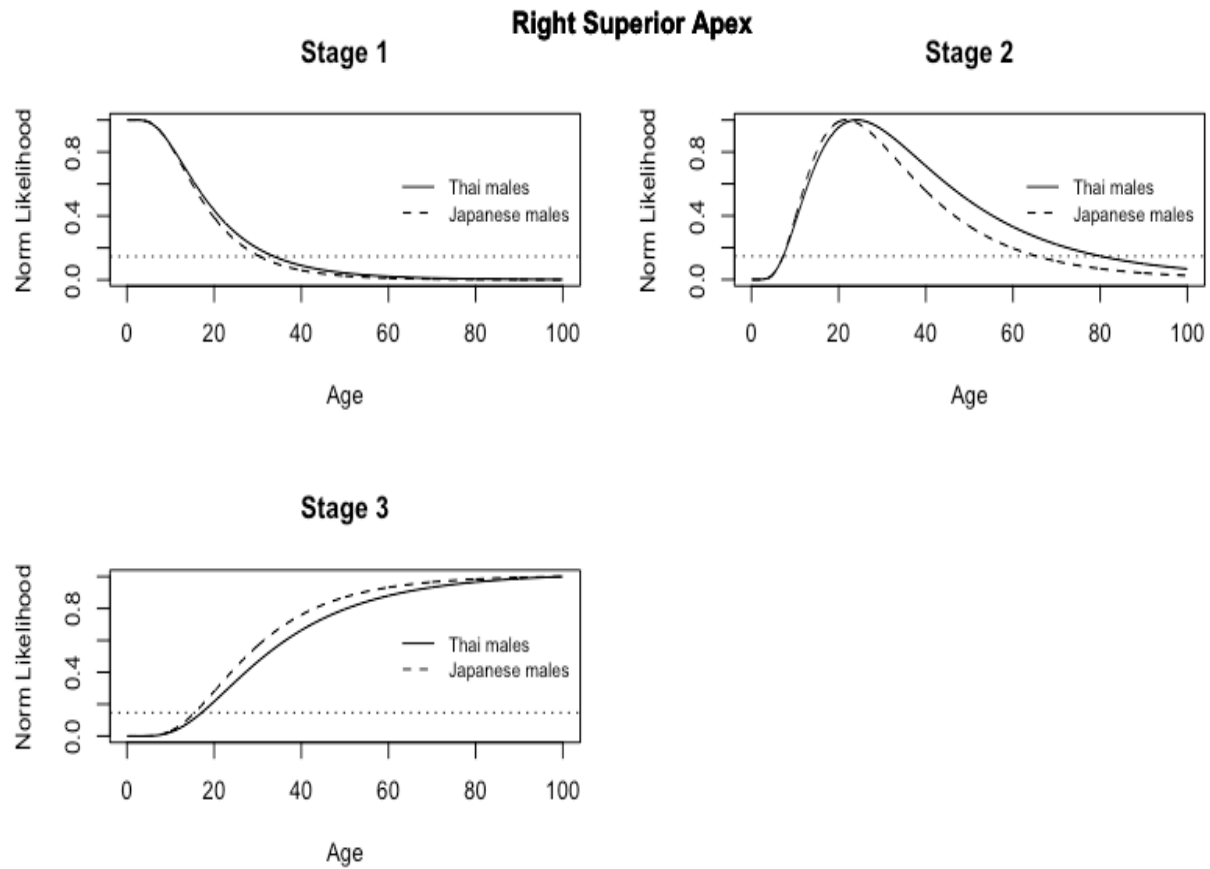


Figure 13. The probabilities of being in each stage of the right superior apex: Japanese and Thai male comparisons

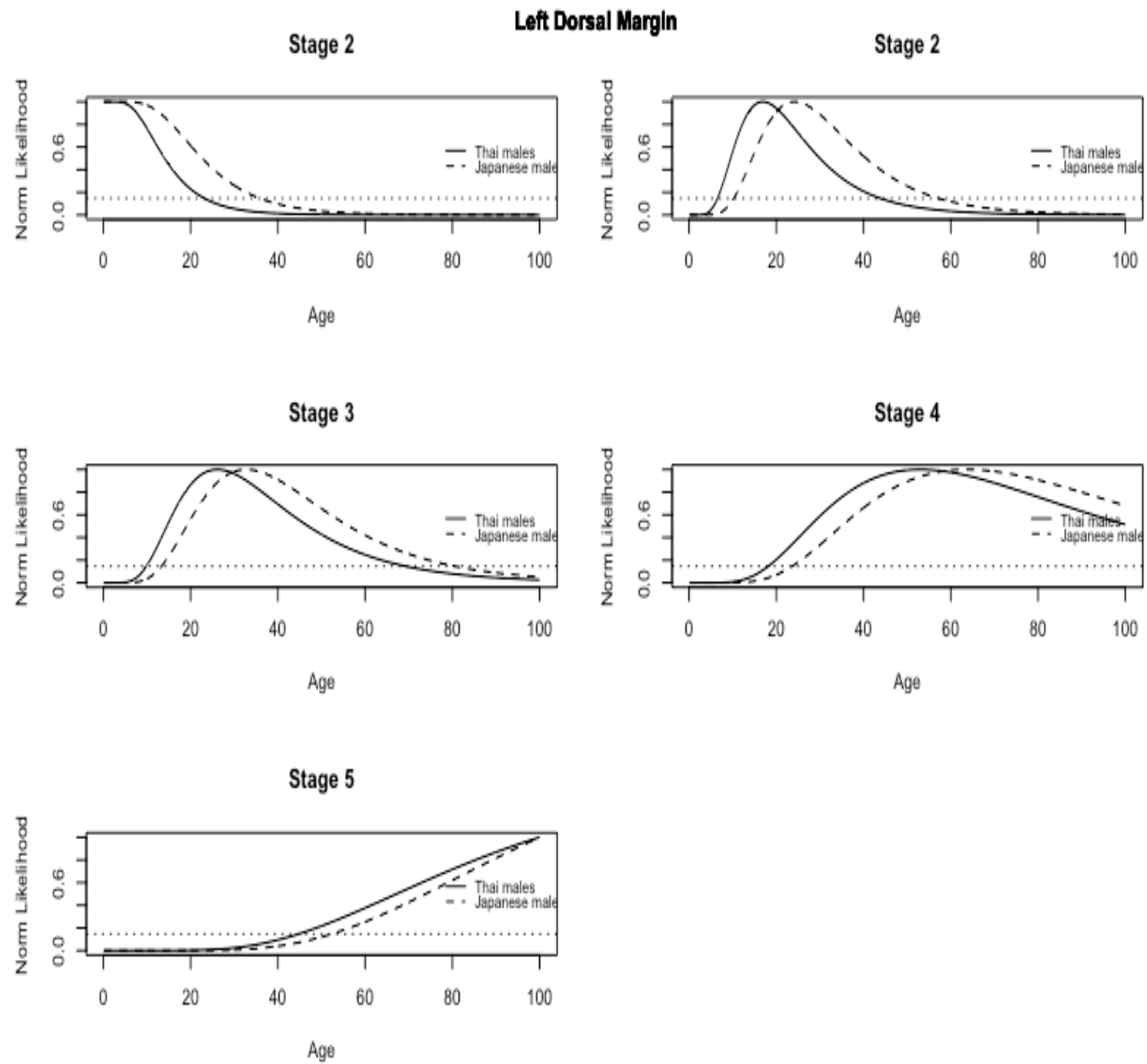


Figure 14. The probabilities of being in each stage of the left dorsal margin: Japanese and Thai male comparisons

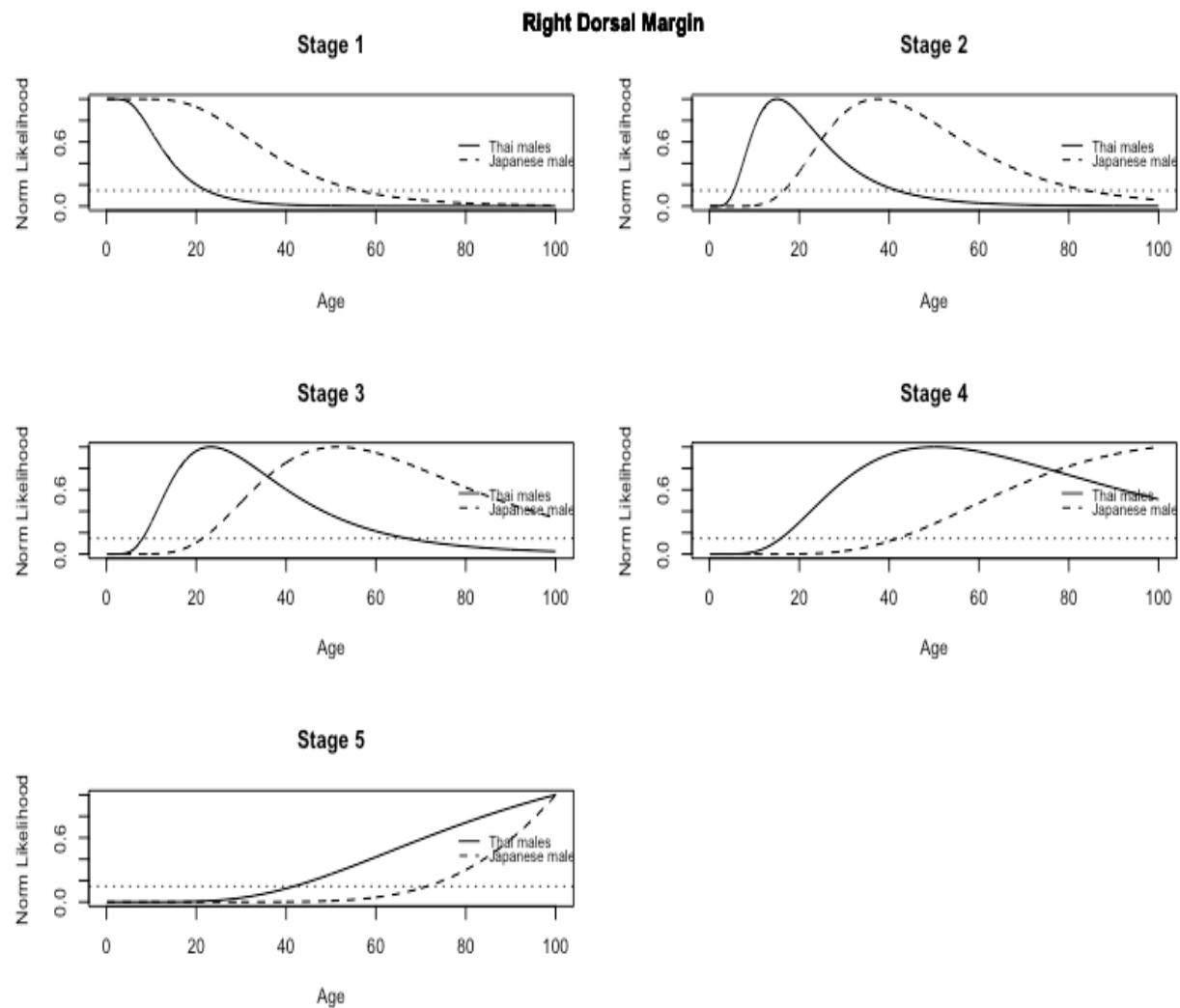


Figure 15. The probabilities of being in each stage of the right dorsal margin: Japanese and Thai male comparisons

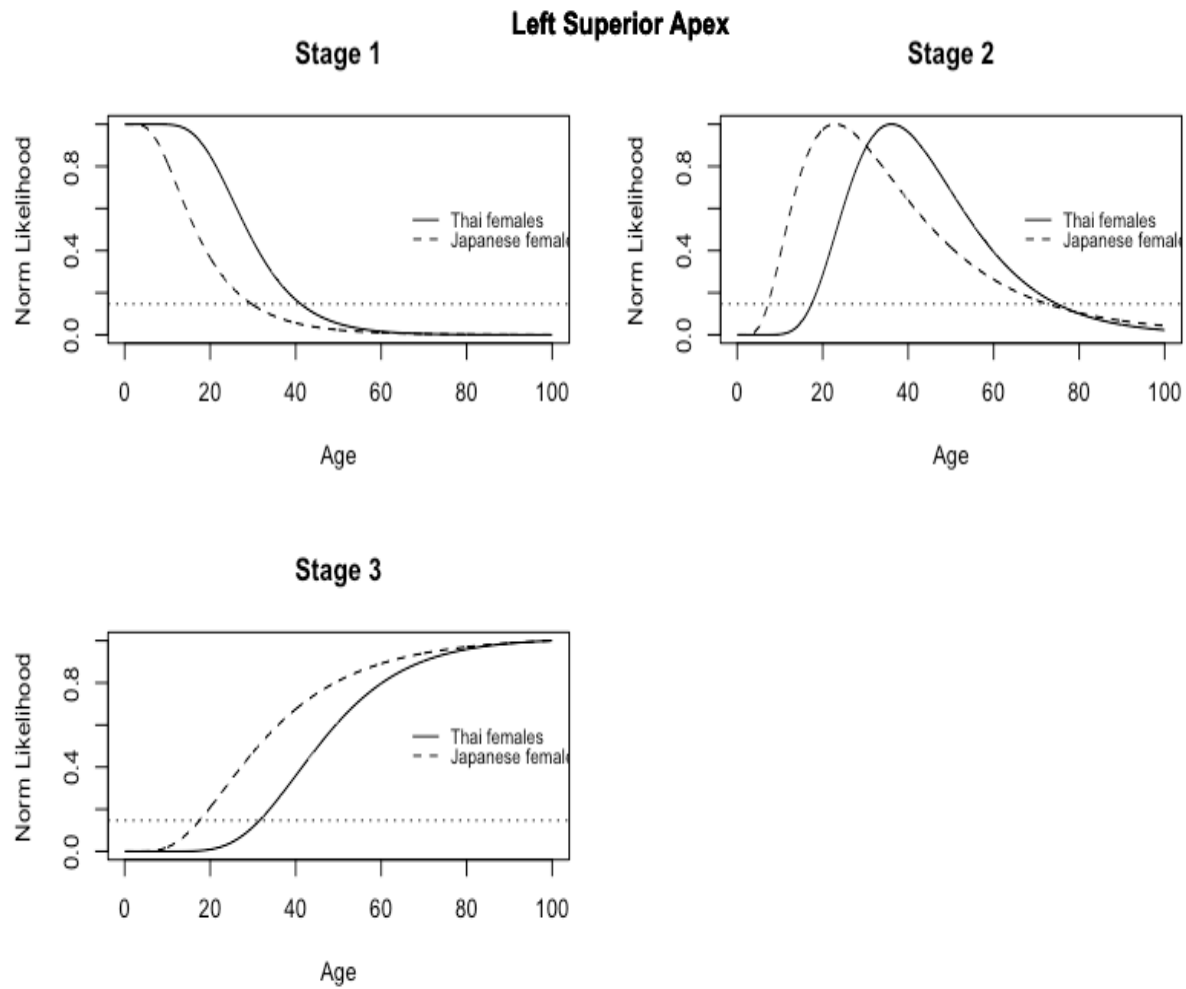


Figure 16. The probabilities of being in each stage of the left superior apex: Japanese and Thai female comparisons

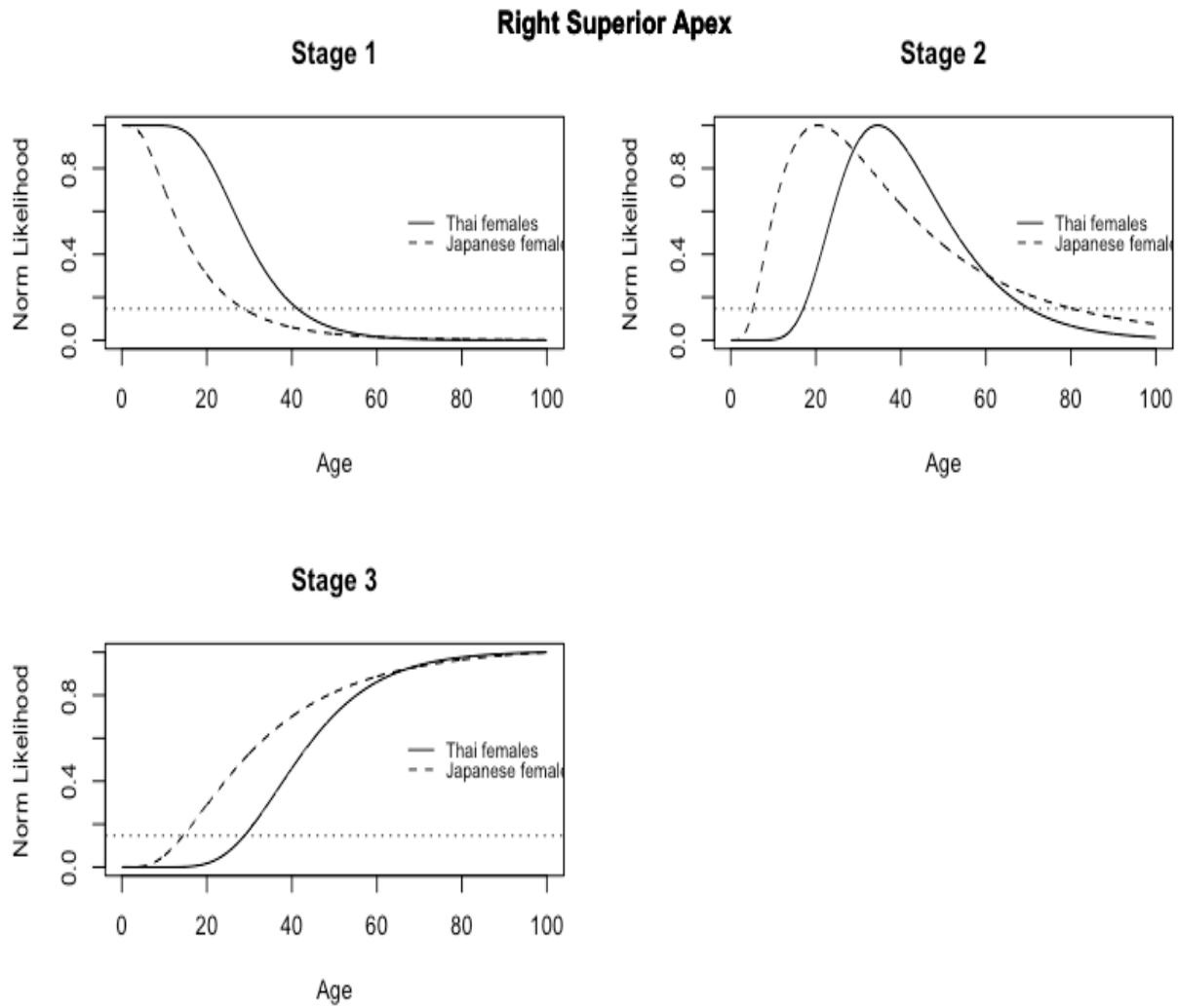


Figure 17. The probabilities of being in each stage of the right superior apex: Japanese and Thai female comparisons

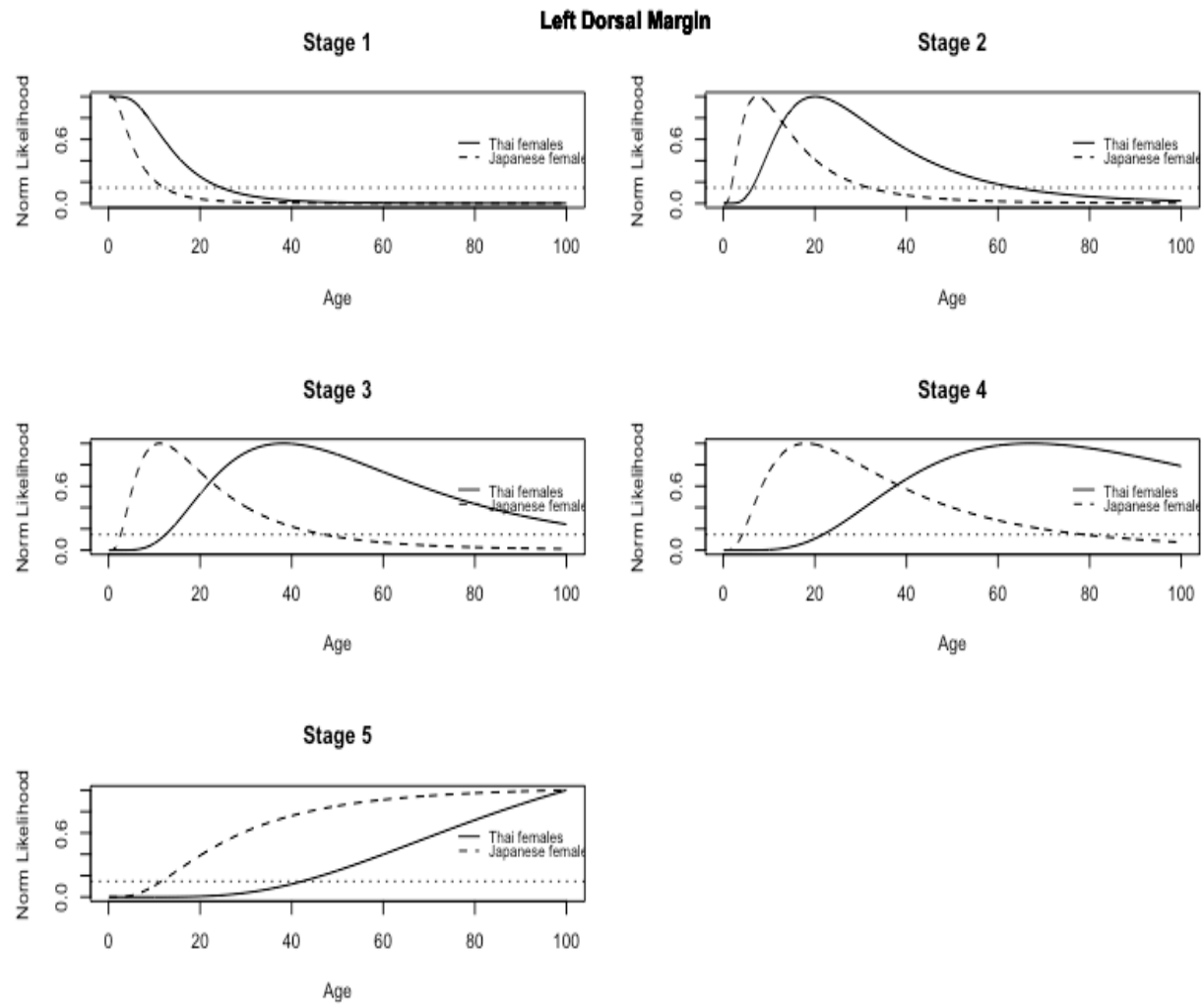


Figure 18. The probabilities of being in each stage of the left dorsal margin: Japanese and Thai female comparisons

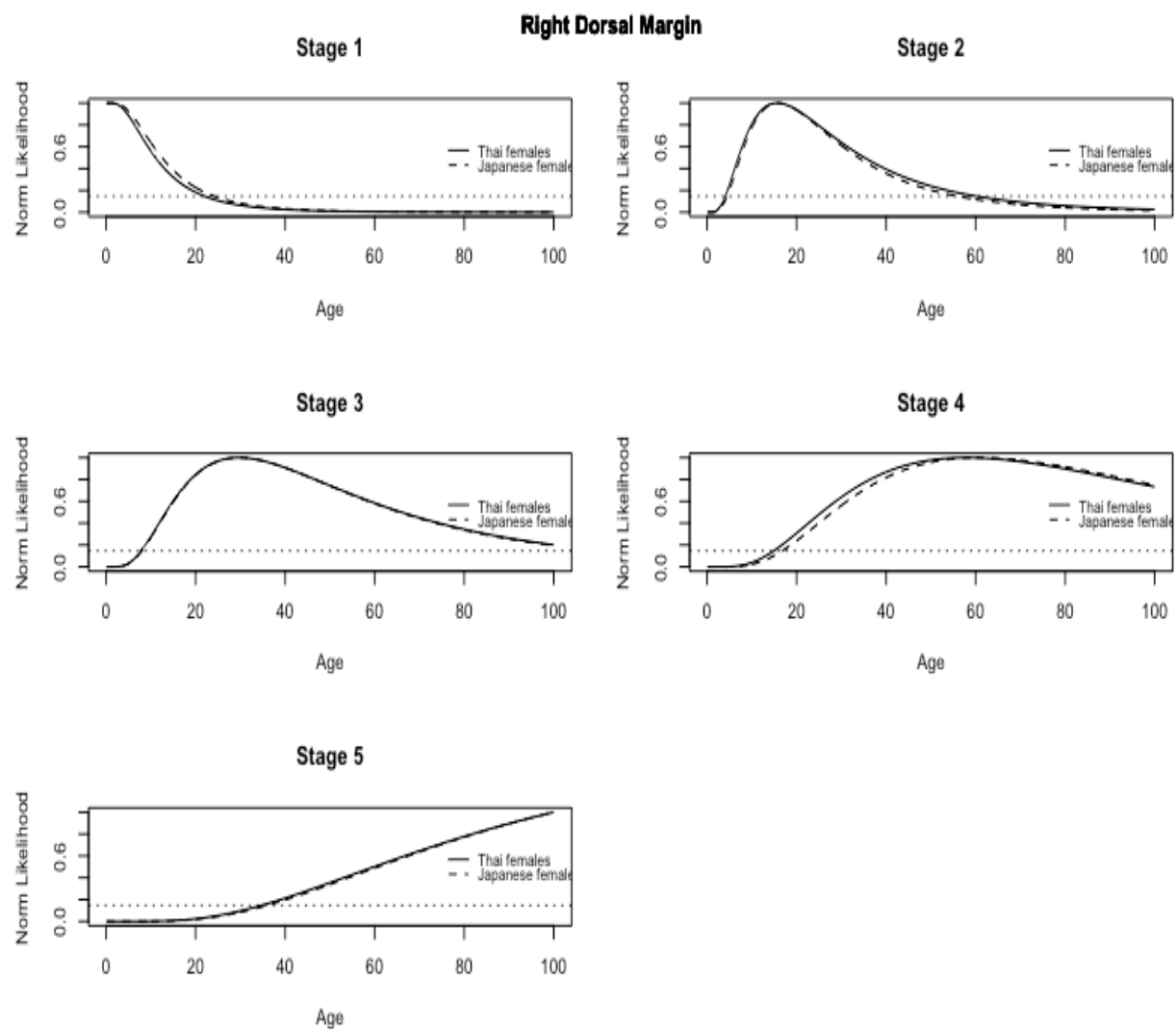


Figure 19. The probabilities of being in each stage of the right dorsal margin: Japanese and Thai female comparisons

Quadrivariate analysis

Although the results of a univariate analysis are somewhat unclear to confirm between-population differences, the results of quadrivariate analysis suggest minimal population differences in skeletal aging, at least within the Asian samples in this study. Table 16 presents the most probable combination of stages that we expect to see for those 20-90 years of age with a 10-year increment. For example, for Japanese males and Thai males who died at age 20, we are expected to observe Stage 2 for left and right superior apices and the left dorsal margin, and Stage 3 for the right dorsal margin. Interestingly, a suit of traits that are likely to be observed for each age is notably consistent between groups—although Thai females at age 20 seems to be the only one that show a quite different combination of traits (Stage 1 for left and right superior apices and Stage 2 for the left and right dorsal margins).

Further corroborating a possible lack of population-specificity in skeletal aging, Figures 20-21 show posterior distributions of individual traits (dashed line) as well as a posterior distribution of all four traits combined (bold solid line) of a Japanese male and Thai male with the same suite of scores. In these two figures, both Japanese and Thai males have scores of left superior apex in Stage 2, while the rest of the three traits are in Stage 3 (i.e. 2, 3, 3, 3.). The abbreviations, AL and AR, stand for left and right superior apices, and DL and DR stand for left and right dorsal margins. It is evident that the two posterior distributions of the four traits combined cover almost identical age ranges, and also the two groups share very similar point estimates: 27.1 years for Japanese males and 28.7 years for Thai males. Note that the set of score “2, 3, 3, 3” is what we expect to observe given an exact age of 30 years for both Japanese and Thai males. This is the likelihood estimate in terms of the Bayes’ theorem. On the other hand, the final age estimates of 27 years and 29 years are the point estimates of highest posterior

probabilities after incorporating a prior. Therefore, one should not expect the posterior probabilities to yield the same age as the likelihood estimates (30 years).

The above results further raise an additional question: how influential one or two trait differences are to final age estimates. Will a Japanese female with trait scores of “3, 3, 4, 4” have different age estimates than a Thai female with traits scores of “3, 3, 3, 4”? Figure 22-23 provide some insight. The two sets of scores with one trait difference still yielded quite similar age estimates: 59 years for Japanese females and 61.6 years for Thai females. The results of the quadrivariate probit analysis show that two geographically distributed populations, and yet from the same continent, not only share similar age-related characteristics, but also the timing of skeletal aging is not much different. Thus, it demonstrates that the multivariate ordered probit regression combined with a Bayesian approach effectively handles variable trait expressions (or outlying scores) observed in two different skeletons of the same age.

Another interesting trend to note is that a certain combination of traits, namely Stage 3 for left and right superior apices and Stage 4 for the left and right dorsal margins (i.e. 3, 3, 4, 4), persist from age of 40 up to 70-80 years for male groups. This indicates a lack of changes in age-related trait expressions for almost 30-40 years for males. If age estimation is made based on these particular four traits, a wide age range with a lack of precise age estimates will naturally result. This suggests why it is important to incorporate as many traits as possible, and find skeletal traits that are exclusive to a certain age cohort to improve precision. In females, however, such trend of prolonged characteristics was less extreme, suggesting females experience greater variability in age-related trait expressions than males.

Table 16. Most probable scores given an exact age between 20-90 years. The four scores are in order of left and right superior apices, and left and right dorsal margins

Age	Japanese males	Thai males	Japanese females	Thai females
20	2 2 2 3	2 2 2 3	2 2 2 3	1 1 2 2
30	2 3 3 3	2 3 3 3	2 2 3 3	2 2 2 3
40	3 3 4 4	3 3 4 4	3 3 4 3	2 2 3 3
50	3 3 4 4	3 3 4 4	3 3 4 4	3 3 3 4
60	3 3 4 4	3 3 4 4	3 3 4 4	3 3 4 4
70	3 3 4 4	3 3 4 4	3 3 4 5	3 3 4 5
80	3 3 5 5	3 3 4 4	3 3 5 5	3 3 5 5
90	3 3 5 5	3 3 5 5	3 3 5 5	3 3 5 5

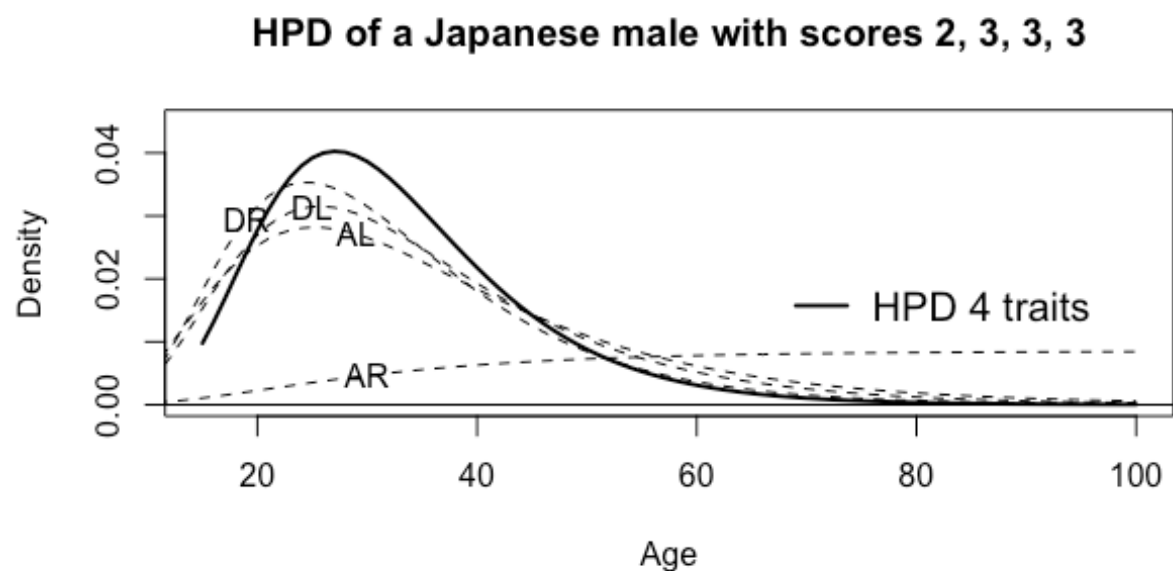


Figure 20. Posterior distributions of a Japanese male with a combination of stages 2, 3, 3, 3

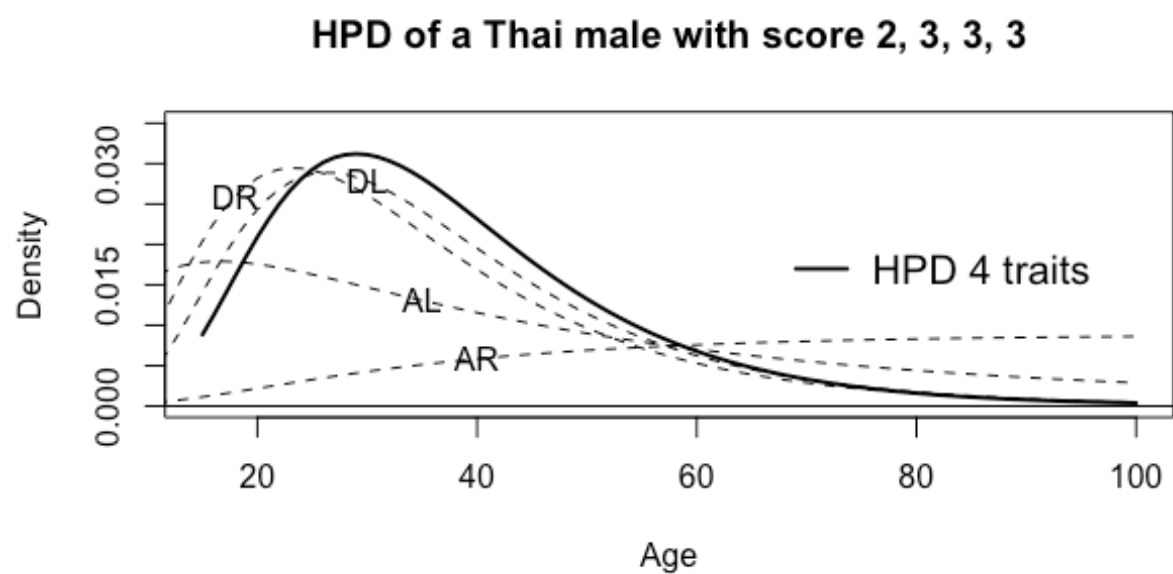


Figure 21. Posterior distributions of a Thai male with a combination of stages 2, 3, 3, 3

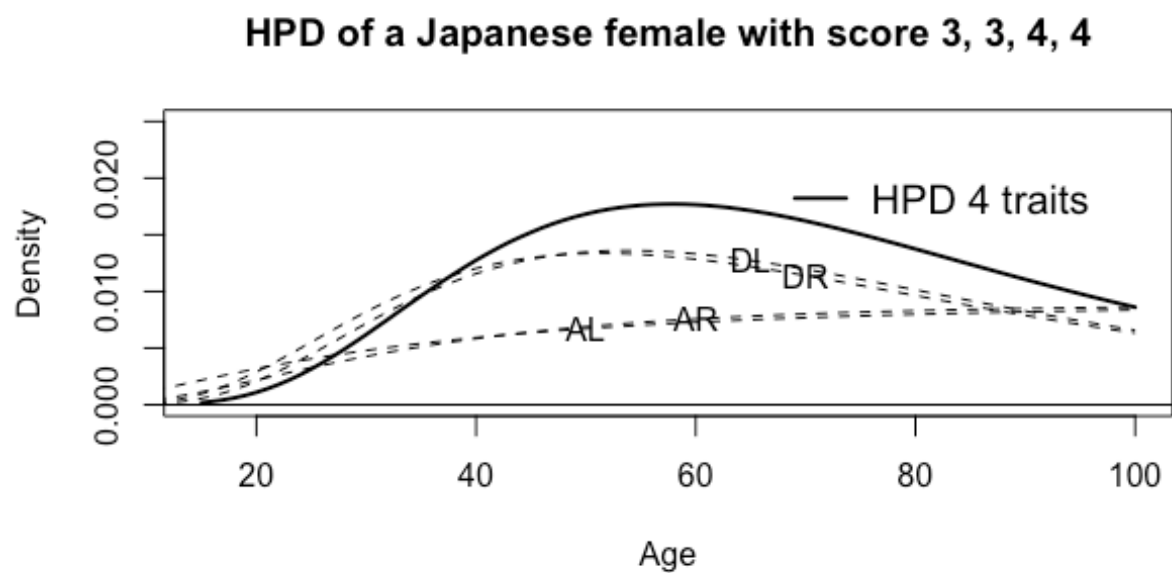


Figure 22. Posterior distributions of a Japanese female with a combination of scores 3, 3, 4, 4

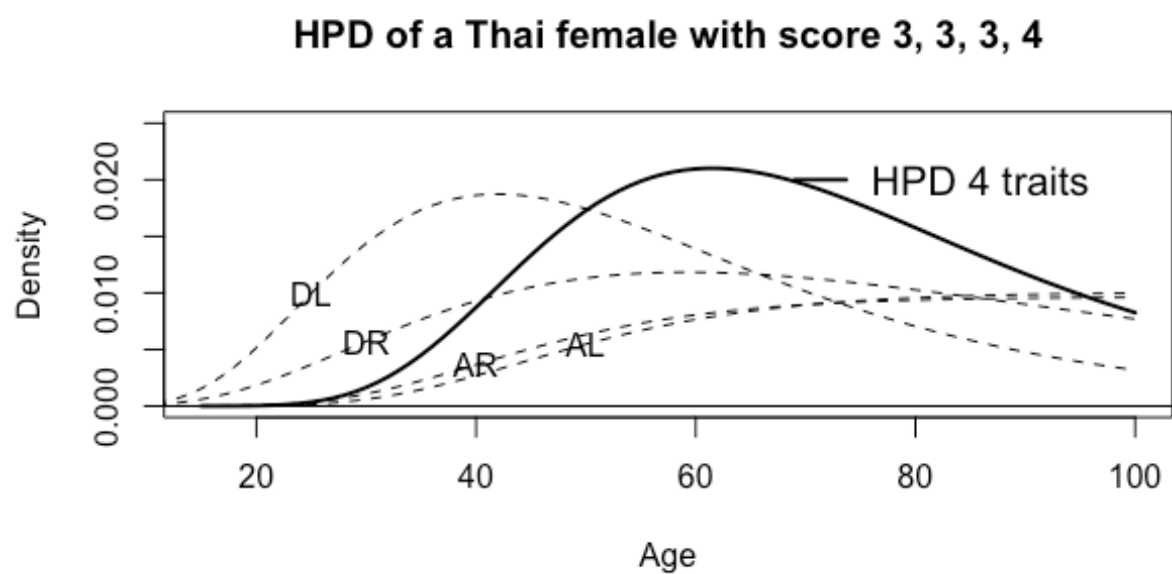


Figure 23. Posterior distributions of a Thai female with a combination of scores 3, 3, 3, 4

Additional comments on side differences

In addition, the results of the analysis suggested that there is no significant side difference in aging rates between sex- and population-specific Asian samples. Although there is a slight side difference in mean ages at transition, posterior probabilities of being in each stage showed that age estimates of left and right sides are almost identical. Figures 20-23 display posterior probabilities of left and right superior apices (AL and AR) and left and right dorsal margins (DL and DR) in dashed line. It is clear that left and right superior apices in Stage 3 (Figures 22-23) as well as left and right dorsal margins in Stage 4 (Figures 20-21) yield considerably overlapping posterior distributions.

Multivariate analysis: Japanese male age estimation using Thai male parameters

The univariate and quadrivariate probit regression above presented age estimation using parameters from the same population. As a last step of evaluating the effects of different aging rates between Thais and Japanese to final age, cross-validation was conducted using the sex-and population-specific probit regression parameters from the opposite population. This part of analysis will help answer how accurately one age estimation model developed on a particular group performs when it is applied to another group.

A parametric Gompertz hazard model was fitted to the Japanese mortality data ($n=8,588,512$) as well as the documented chronological ages of the Japanese collections ($n=285$). Given the youngest male in the Japanese sample is 17 years old, the Gompertz function was adjusted to begin from 17 years. Two Gompertz parameters, namely $a_3=0.0002$ and $b_3=0.1037$, were obtained from the Japanese male mortality datab, while Gompertz parameters of the collection ages are $a_3=0.0065$, and $b_3=0.0508$. As Figure 24 illustrates, the two mortality profiles

are considerably dissimilar, and, therefore, provide a great opportunity to evaluate effects of different informed priors to final age estimates.

As shown in Table 17, ages at death of middle-aged adults were estimated with the smallest error (7-9 years). Age estimates for both young and old Japanese adults showed a similar magnitude of error: 20-22 years for young Japanese adults, and 20-21.5 years for old Japanese adults. Unfortunately, using two different informed priors did not dramatically change the age estimates as hoped, although there was slight improvement in age estimates based on the collection-age prior for young and middle-aged Japanese males (a difference between RMSE values was only 1 year, Table 17).

Multivariate analysis: Thai male age estimation using Japanese male parameters

Ages at death of Thai males (n=157) were estimated based on the probit parameters of the Japanese male sample to assess the utility of Japanese parameters to Thai males. For Thai male age estimation, probit regression parameters from 11 traits of the Japanese male sample were used: left and right superior apices, ventral margins, dorsal margins, superior and inferior demiface topography of the sacroiliac joint, and interpalatine sutures. For Thai males, an informed prior was obtained by fitting a Gompertz hazard model to documented male ages of the two Thai collections (n=488). The documented male ages yielded the following two Gompertz parameters, $a_3=0.0001$, $b_3=0.1043$. As the youngest male in the Thai male sample was 14 years old, the Gompertz function was shifted to start from 14 years.

Table 18 shows that, for Thai males, Japanese parameters yielded the lowest error for the middle-aged adults (8-9 years). This is the same result already observed in Japanese males. The only differences in RMSE trends between Japanese and Thai males are that, while young and old

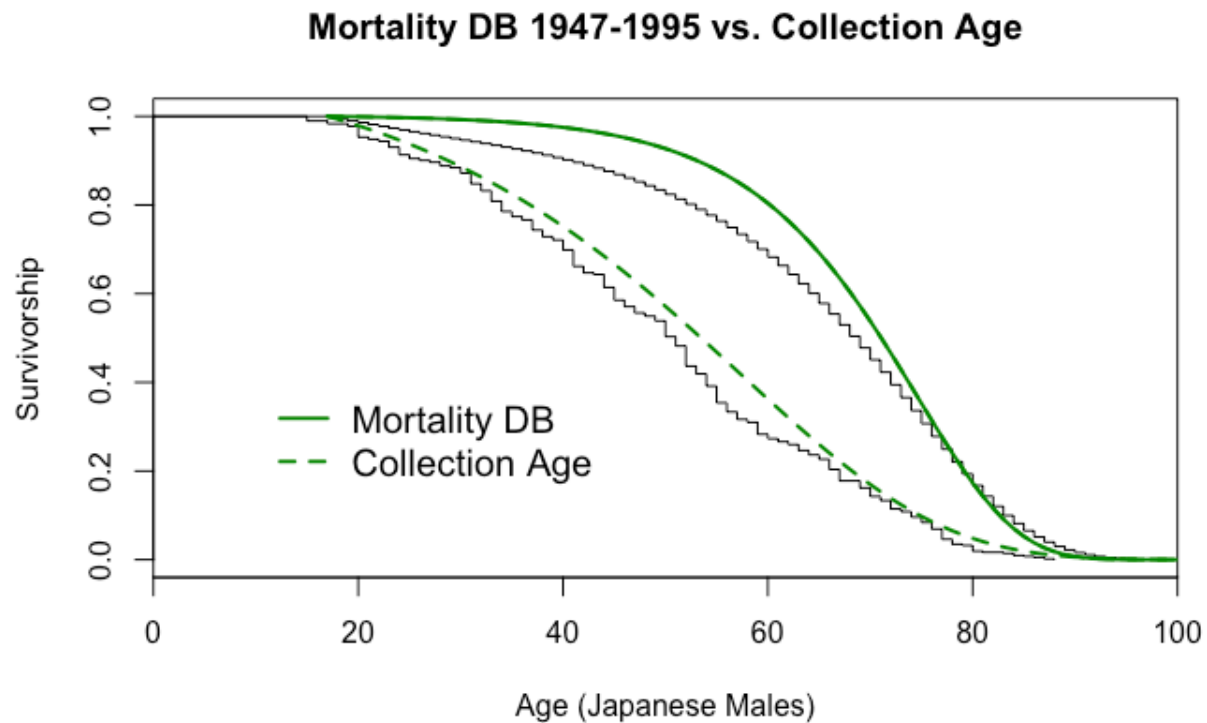


Figure 24. Two Gompertz hazard models for Japanese males. A solid line is for 1947-1995 mortality database (DB) of Japanese males. A dashed line is for documented chronological ages of the Japanese collections

Table 17. Japanese Male RMSE: Thai male parameters based on 4 traits

	Young (<40yrs) n=40	Middle (41-60yrs) n=82	Old (61yrs<) n=16
Gompertz Prior 1 (Mortality DB)	21.92	8.09	20.22
Gompertz Prior 2 (Collection Age)	20.45	6.82	20.32
Uniform Prior	21.16	8.75	21.59

Table 18. Thai Male RMSE: Japanese parameters based on 11 traits

	Young (<40yrs) n= 11	Middle (41-60yrs) n=45	Old (61<yrs) n=101
Gompertz Prior	22.53	7.85	17.43
Uniform Prior	24.88	8.77	15.99

Japanese adults had the same magnitude of error around 20 years, young and old Thai males showed a different magnitude of error: Young Thai males had greater RMSE values (22.5-25 years) than older Thai males (16-17 years).

Why do we see such a trend in RMSE values of Japanese males and Thai males?

The consistent trend in RMSE values—middle-age adults with the lowest error—observed in Japanese and Thai males can be understood by observing the distribution of age estimates.

Figures 25-26 show point age estimates, 95% age ranges for Japanese and Thai males in comparison to their documented chronological ages, and a vertical line is an estimated age range for each individual. Filled circles are documented ages, while a filled circle with an “x” marker is a point age estimate. In Figure 25, point estimates are heavily concentrated on the middle-age range from 40-60 years for all age groups. This explains why the middle-aged Japanese males had the most accurate age estimates, while the young and old showed increased inaccuracy. Likewise, the pattern of age estimates for Thai males (Figure 26) shows homogenous point estimates around the middle-aged category. However, in comparison to Japanese males, Thai male point estimates are shifted approximately 10 years upward, and so concentrated around 50-70 years. The homogenous point estimates further resulted in overestimation of the young and underestimation of the old (Figure 27) with increased RMSE values in young and old age groups, and created a unique U-shaped (squared) error distribution for both Japanese males and Thai males (Figures 28-19). The over-and underestimation seemed to be stronger than ADBOU age estimates (Figure 5), and persisted in estimates of other groups throughout the analysis.

Another tendency to note in the above RMSE values is the notably reduced error for old Thai males compared to the Japanese counterpart. Other than the fact that the probit parameters

are from the opposing population, the only difference—which may have contributed to reducing RMSE values—is a number of traits used for age estimation. Note that, for Japanese males, Thai male probit regression parameters of four traits were used, whereas, for Thai males, Japanese male parameters derived from eleven traits are used for age estimation. This suggests that adding seven more traits to age estimation favorably improved age estimates: 16-17 years for old Thai males as opposed to 20-21.5 years for old Japanese males. Nevertheless, one might still question effects of different populations for causing different RMSE values. Would using probit regression parameters from the same population improve age estimates? A short answer is ‘no.’ Concerning this question, age estimation using multivariate probit regression parameters from the same population was conducted, but there was no notable improvement in age estimates: For Japanese males, the RMSE values were 18.5 years for the young, 10 years for the middle-aged, and 15 years for the old. For Thai males, the RMSE values were 24 years for the young, 9 years for the middle-aged, and 19 years for the old. Ironically, for old Thai males, the RMSE value was even bigger when Thai parameters were used than using Japanese parameters.

For young males, however, RMSE values remained high for both Japanese and Thais (20-25 years). This suggests that, in the current data, skeletal traits representing young individuals were poorly captured. This should not be mistaken that the current TA method does not perform well for young individuals. In fact, ADBOU excelled in young Asian adult age estimation (Tables 9-10, Figures 3-4) than the current population-specific model. Lastly, although some improvement in point age estimates was detected after incorporating a more number of traits, there was no noticeable improvement of estimated age ranges. An average width of estimated age ranges was 63 years for Japanese males and 70 years for Thai males

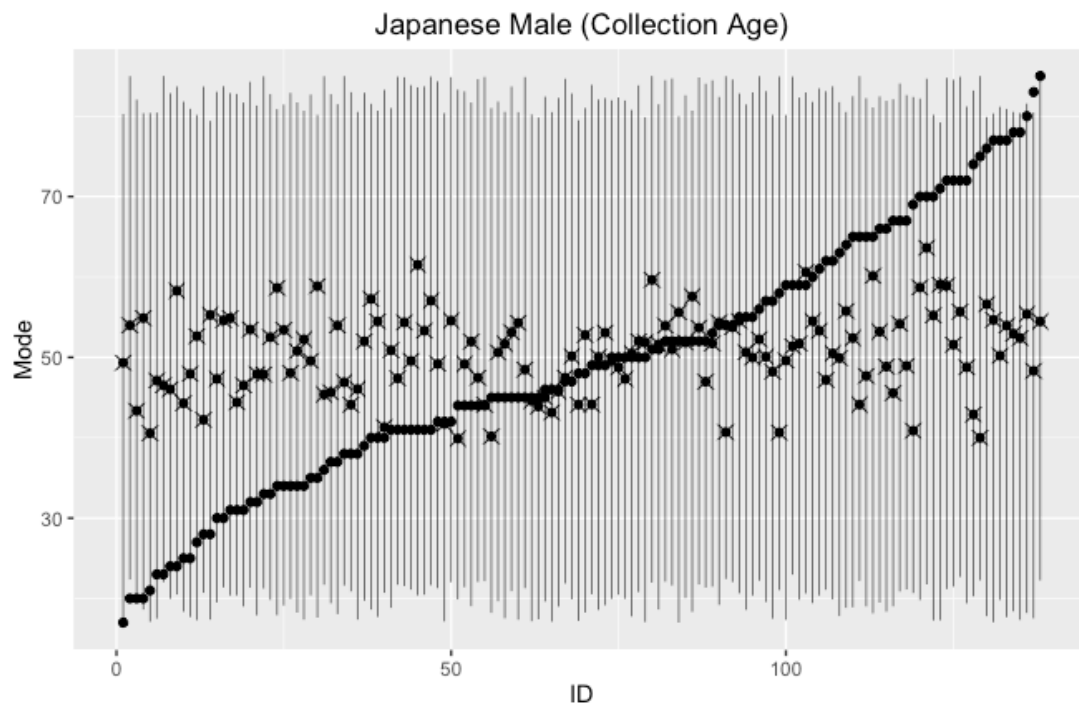


Figure 25. A distribution of estimated age ranges, point estimates, and documented chronological ages for Japanese males

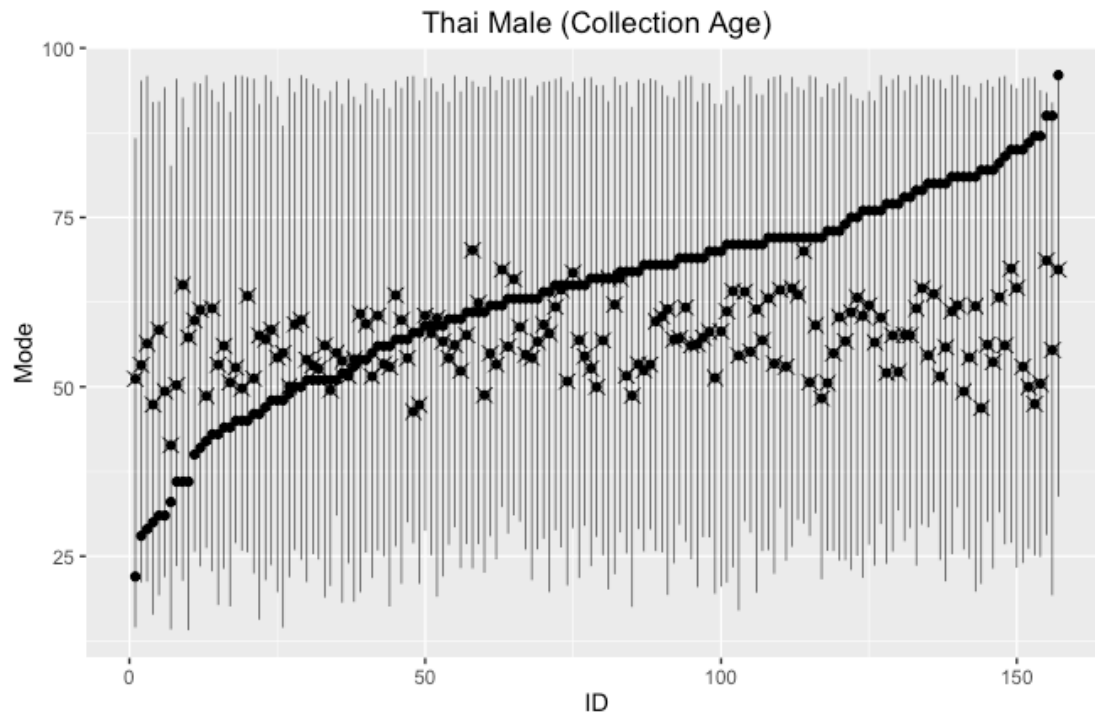


Figure 26. A distribution of estimated age ranges, point estimates, and documented chronological ages for Thai males

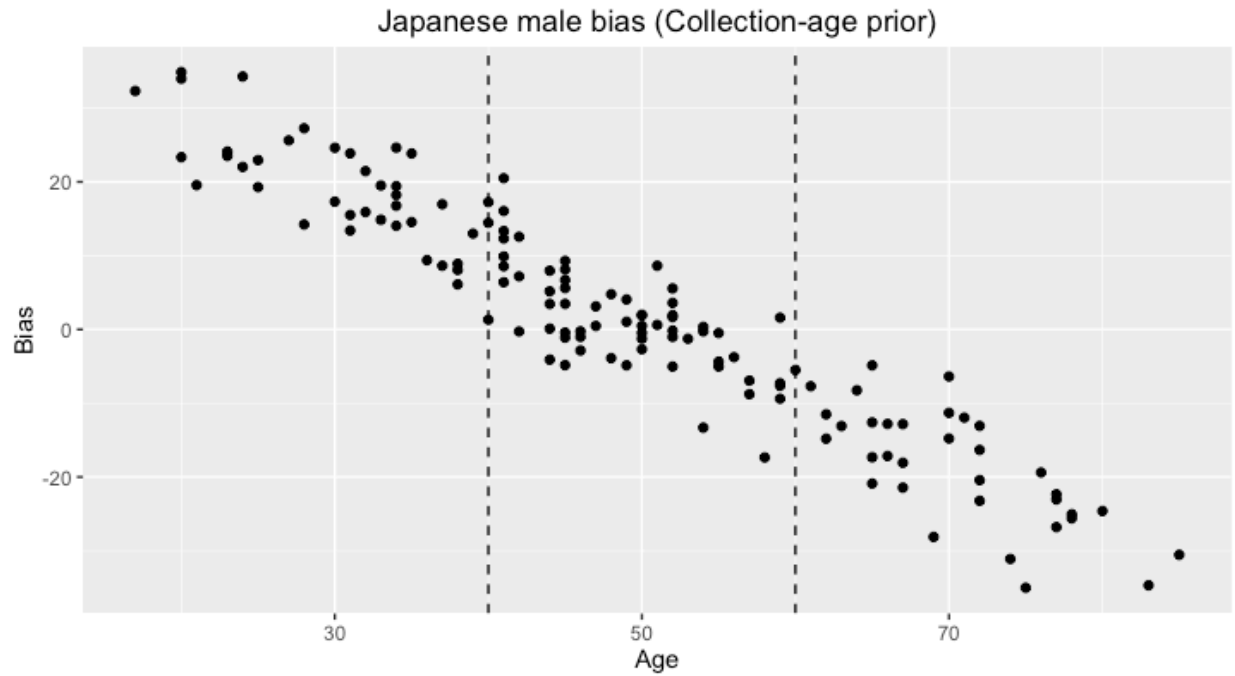


Figure 27. Bias of Japanese male age estimates using Thai male probit parameters

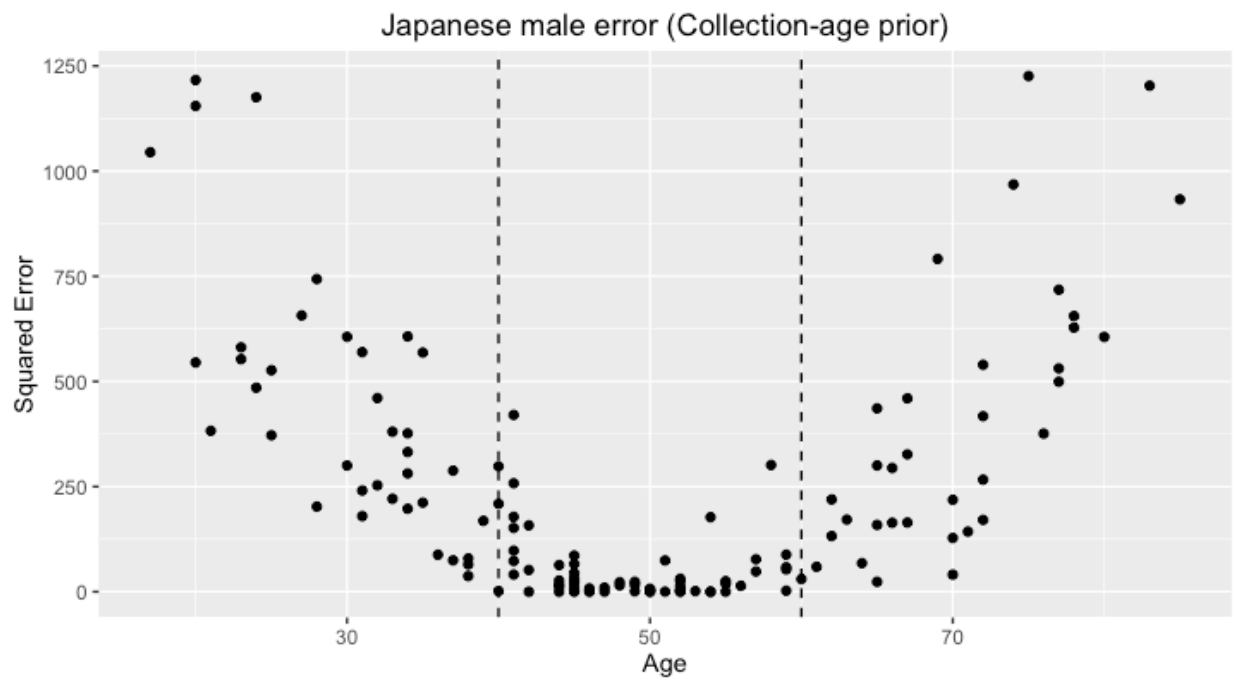


Figure 28. Japanese male squared-error distribution

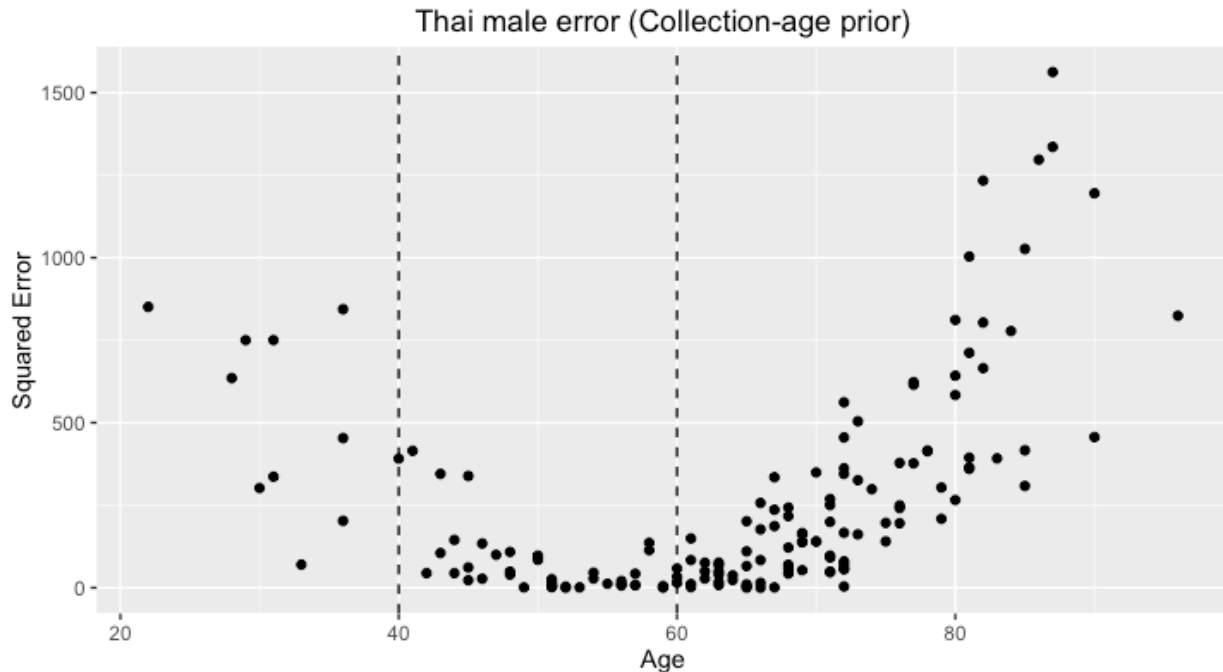


Figure 29. Thai male squared-error distribution

(Table 27). As Figures 25 and 26 show, estimated age ranges cover almost entire adulthood, providing very little information if one needs more precise estimates to narrow down possible matches, such as in forensic cases.

Multivariate analysis: Japanese female age estimation using Thai female parameters

TA scores of the Japanese female sample (n=45) were analyzed based on Thai female probit parameters to obtain age estimates. Eight skeletal traits of the Thai female sample were used to obtain multivariate probit regression parameters. The traits include left and right superior apices, ventral margins, and dorsal margins, and left superior and inferior demiface topography. As it was the case for the Japanese male sample, two independent informed priors were combined with the probit parameters to obtain final age estimates of Japanese females. The first informed prior is a Gompertz hazard model fitted to the Japanese female mortality data ranging from

1947-1995 ($n=7,459,018$) with Gompertz parameters of $a_3=0.0004$ and $b_3=0.0999$. The second informed prior was created using the documented female ages of the Chiba and Jikei collections ($n=104$), and Gompertz parameters of $a_3=0.0065$ and $b_3=0.0516$ were attained. As the youngest individual in the Japanese female sample are 23 years, the Gompertz survivorship was shifted accordingly. The two distinctively different Gompertz survival functions for Japanese females are presented in Figure 30.

As shown in Table 19, Thai female parameters yielded the most accurate age estimates for middle-aged Japanese females (9-11 years) followed by the old (14-15 years) and the young (24-26 years). This trend in RMSE values is consistent with that of Thai males. The role of two distinctly different priors seemed to be negligible in improving age estimates of Japanese females as RMSE values stayed about the same regardless of using informed and uniform priors.

Multivariate analysis: Thai female age estimation using Japanese female parameters

Ages at death of Thai females ($n=79$) were estimated using the probit parameters attained from four pubic symphyseal traits of Japanese females (left and right superior apices and dorsal margins). A Gompertz model was fitted to the documented female ages of the two Thai collections ($n=252$) with two parameters of $a_3=0.0002$, and $b_3=0.1017$. The Gompertz function was adjusted to reflect the youngest individual (20 years) in the Thai female sample.

RMSE values in Table 20 show a consistent pattern of error that has been observed previously. The most accurate age estimates were made for middle-aged females followed by the old and the young. The RMSE values of young Thai females are the largest of all groups so far (29-33 years). A very small sample size of young individuals ($n=3$) in Thai females may have

exacerbated the error. The old Thai females had larger RMSE values (17-18 years) than those of Japanese elderly females (14-15 years). Any benefit of using an informed prior was not detected.

Consistent patterns of RMSE values for the females groups and the male groups

The Japanese and Thai female samples showed the same trend that was already observed in Japanese and Thai male samples: Middle-aged adults had the lowest error, followed by the old and the young. Thus, the pattern of age estimates for the female groups also paralleled that of the male groups. Figure 31 and Figure 32 plot both age estimates and documented ages along with estimated age ranges. Both Japanese and Thai females have point estimates that are heavily concentrated within the 50-70 year range. This pattern of point estimates resulted in the lowest error for middle-aged females, while increasing RMSE values for young and old females. As a result, extensive overestimation of the young and underestimation of the old persisted for the female groups (Figure 33).

As it was the case for the male group, it is worth paying attention to the differences in the RMSE values between Japanese and Thai females. Particularly, the RMSE values for the young and old Japanese females are smaller than those of Thai female counterparts. The reduced error seems to be attributable to the larger number of traits used for Japanese female age estimation: Age estimates of Japanese females were derived from eleven traits, while only were four traits used for Thai female age estimation. Consistent to the results of the male groups analysis, adding a more number traits favorably improve age estimates for females: 14-15 years for old Japanese females as opposed to 17-18 years for old Thai females, and 24-26 years for young Japanese females compared to 29.5-33 years for young Thai females.

Nevertheless, regardless of the number of traits added to the age estimates and a type of

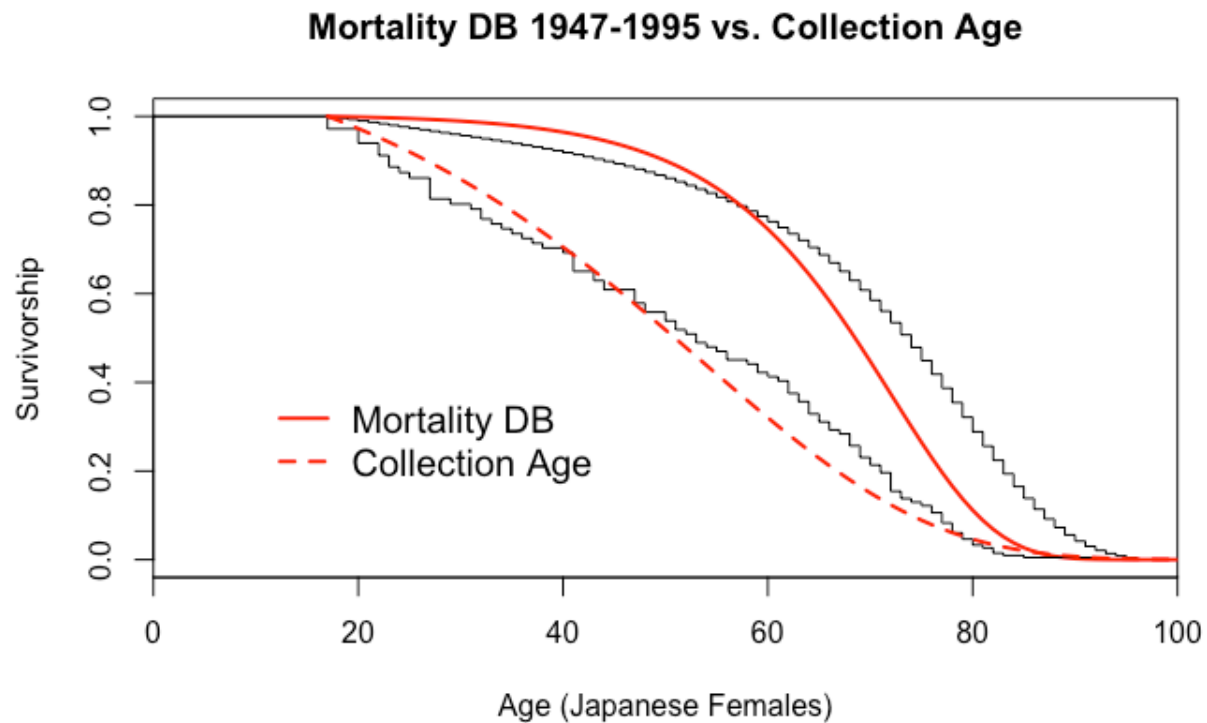


Figure 30. Two Gompertz hazard models for Japanese females. A solid line is for 1947-1995 mortality database (DB) of Japanese females. A dashed line is for documented chronological ages of the Japanese collections

Table 19. Japanese female RMSE: Thai female parameters based on 8 traits

	Young (<40yrs) n=9	Middle (41-70yrs) n=11	Old (71yrs<) n=25
Gompertz Prior 1 (Mortality DB)	26.19	10.96	14.89
Gompertz Prior 2 (Collection Age)	25.58	11.47	14.42
Uniform Prior	23.95	9.17	14.36

Table 20. Thai female RMSE: Japanese female parameters based on 4 traits

	Young (<40yrs) n=3	Middle (41-70yrs) n=22	Old (71<yrs) n=54
Gompertz Prior	29.5395	9.874421	18.46218
Uniform Prior	33.42651	10.13132	17.35603

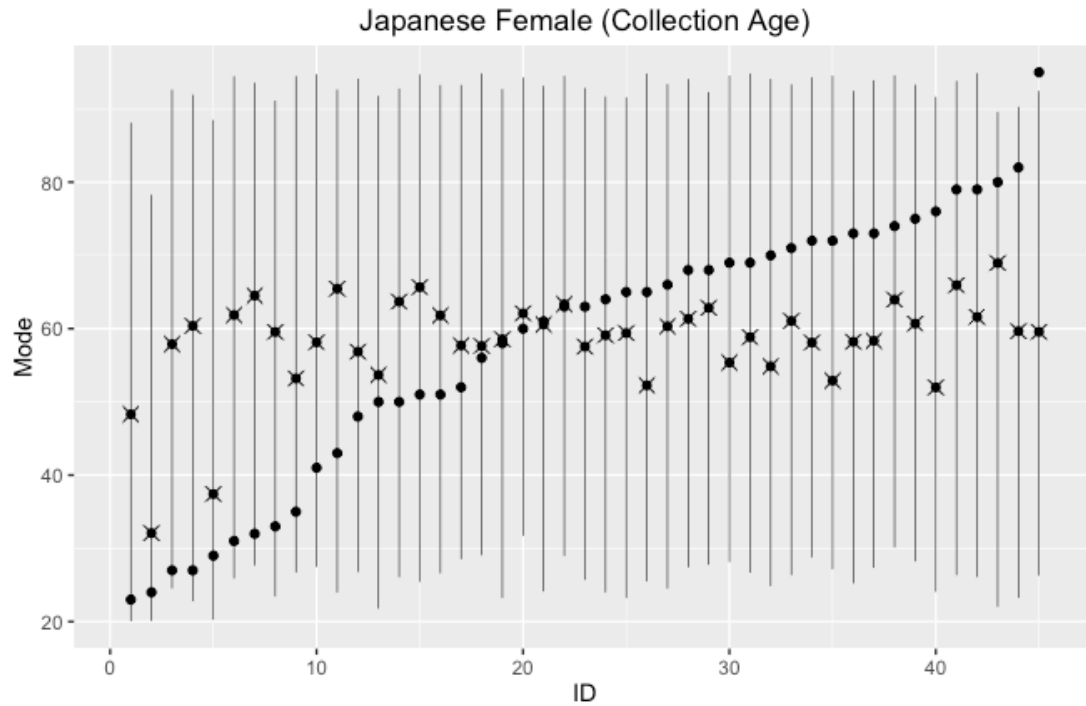


Figure 31. A distribution of estimated age ranges, point estimates, and documented chronological ages for Japanese females

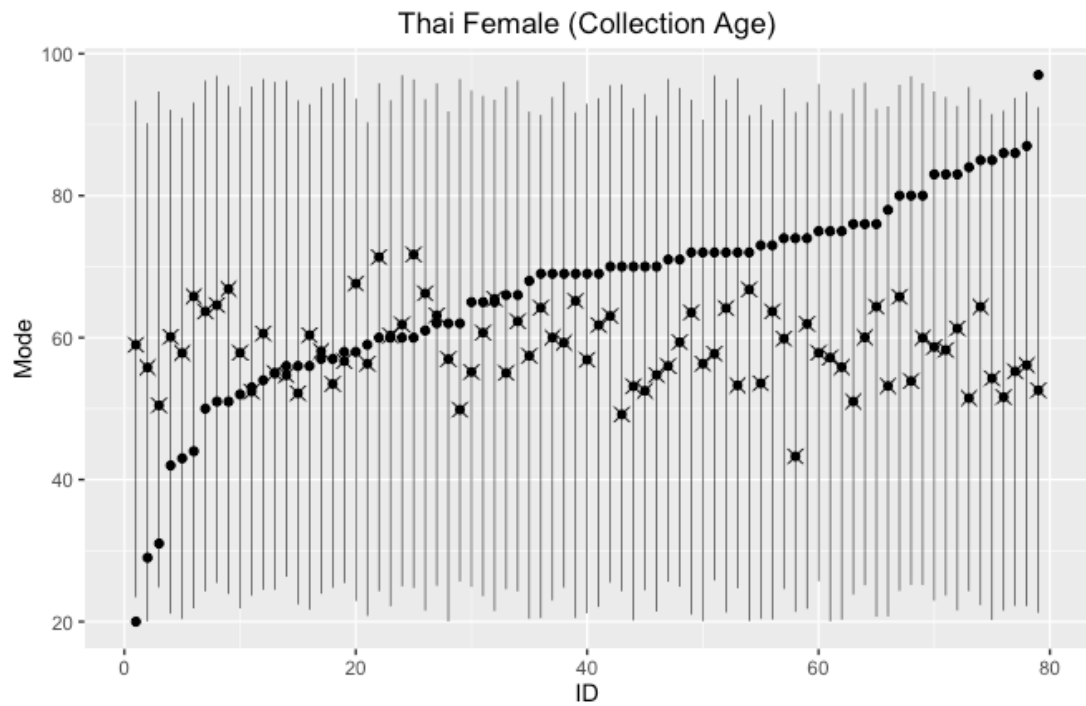


Figure 32. A distribution of estimated age ranges, point estimates, and documented chronological ages for Thai females

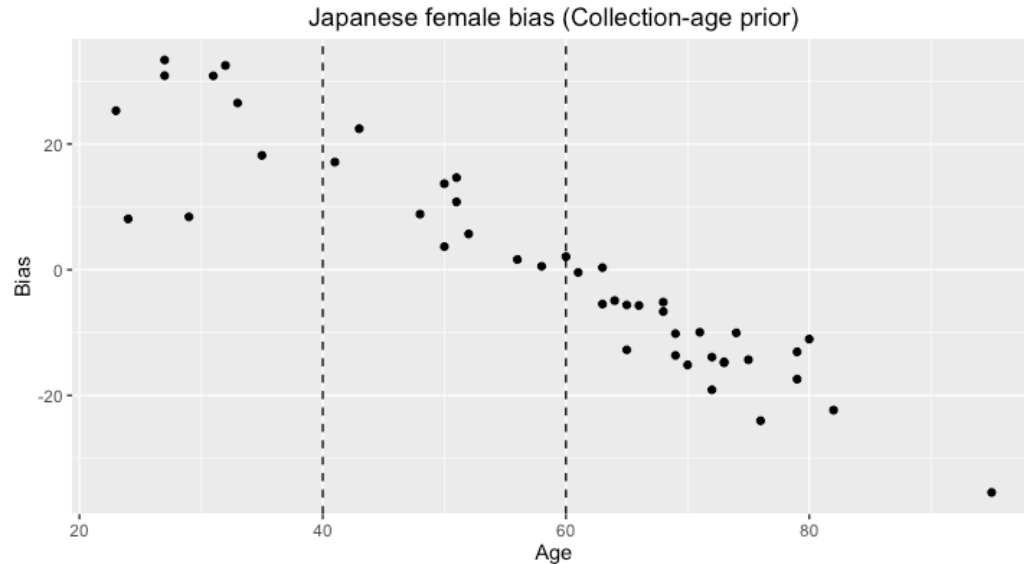


Figure 33. Bias of Japanese female age estimates using Thai female probit parameters

informed priors combined, estimated age ranges remained very wide for both Japanese and Thai males and females (Figures 31-32). However, the average width of age ranges for Japanese females was narrower (67 years) than that of Thai females (71 years), suggesting possible benefits of adding extra traits to age estimates to narrow down age ranges.

Hypothesis (4) Evaluating pooled-sample parameters

The last hypothesis is that *age estimation models based on pooled Japanese and Thai samples will perform better than Japanese- and Thai-specific models and ADBOU or conventional age estimation methods by generating more accurate and unbiased age estimates for the Asian samples*. In order to assess any improvement in age estimates when pooled samples were used, multivariate probit analysis parameters were obtained from pooled Asian males (n=295) and females (n=124). Following Konigsberg's (2015) protocols, performance of the two new pooled-sample probit models was evaluated based on simulated new target samples (n=200).

The Lagrange multiplier test of goodness-of fit was conducted to evaluate the model fit and the normality assumption, and pseudo- R^2 for each of the 33 traits was calculated to examine strength of relationship between traits and documented ages. Table 21 and Table 23 show the results of the Lagrange multiplier test as well as pseudo- R^2 values. The Lagrange multiplier test indicates that 16 traits for males and 12 traits for females are below the significant level 0.05, suggesting inadequate fit of the model to the data and violation of the normality assumption. Instead of discarding these variables, the stage collapsing approach recommended by Konigsberg et al. (2016) was followed. The idea behind the stage collapsing is that the model fit can be improved when a number of stages for a certain trait, which violates the assumptions, are reduced. For example, the left symphyseal relief with a five-stage system can be adjusted to a four-stage system, or as small as a two-stage system based on the Lagrange multiplier test result, to yield a better model fit. However, for a trait with five stages, there are 14 different ways to collapse two adjacent stages, and the possibilities increase as the number of original stages increase. This makes manual evaluation of all possible stage collapsing extremely inconvenient. Fortunately, R scripts for stage collapsing written by Konigsberg conveniently allow researchers to evaluate fit of the model on all possible combinations of stage collapsing, and provide associated p-values of the Lagrange multiplier test. Continuing the symphyseal relief example, Table 22 shows that the five stages are reduced to 2 stages by collapsing Stages 2, 3 and 4, and the second round of the Lagrange multiplier test in Table 21 indicates the probit model fit to the new collapsed data was significantly improved. Another benefit of the stage collapsing is that it increases a strength of relationship between skeletal traits and known ages: Note that pseudo- R^2 for the left symphyseal relief was merely 0.09 before stage collapsing, but it increased up to 0.544 after reduced into a binary trait (Table 21).

After stage collapsing (as shown in Table 22), a second round of the Lagrange multiplier test was conducted on all 33 variables. The last two columns of Table 21 show that none of the traits now has significant p-values. Note, however, the left superior apex, and left and right posterior inferior exostoses were excluded from pooled male samples as there was no improvement even after the stage collapsing, and their relationship to age was close to zero. For the pooled-female sample, the same procedure as the pooled-male sample was followed. Relevant p-values of the Lagrange multiplier test, pseudo- R^2 and collapsed stages for the female group is presented in Tables 23-24. For the same reason as the male group, left and right superior morphology, and left and right superior and inferior exostoses were excluded for the pooled females analysis. The pseudo- R^2 values in Table 21 and Table 23 indicated that a number of traits have much weaker correlation with chronological age than the traditionally reported values (Table 1 in Chapter 2): For males, the left symphyseal relief had the highest pseudo- R^2 value, 0.544, and right posterior spicules rendered the lowest pseudo- R^2 , 0.001. These values indicate that the 54.4% of the variation seen in the left symphyseal relief can be explained by age. For females, left and right superior apex had the highest value, 0.677 and the left posterior spicules had the lowest value, 0.006.

To obtain age estimates of the new target samples, the pooled-male and pooled-female probit regression parameters were combined with a Gompertz function with the same hypothetical parameters, $a_3 = 0.001$ and $b_3 = 0.1$, used for sample simulation, which has been explained in Chapter 4. Tables 25-26 show RMSE values for simulated males and females. Consistent with what we have already observed in sex- and population-specific analyses, the RMSE values for pooled-sample showed that the middle-aged adults had the lowest error, followed by the old and the young. For young and old males, adding all three age indicators

Table 21. Pooled male Lagrange multiplier test and pseudo-R² before and after stage collapsing

Trait	Before stage collapsing		After stage collapsing	
	Goodness-of-fit	Pseudo-R ²	Goodness-of-fit	Pseudo-R ²
PS Relief L	<0.05*	0.09	0.0829	0.544
PS Relief R	<0.05*	0.118	0.7593	0.308
PS Texture L	<0.05*	0.226	0.4906	0.36
PS Texture R	<0.05*	0.172	0.0944	0.096
PS Sup Apex L	<0.05*	0.194	NA	NA
PS Sup Apex R	0.2291	0.196	0.2291	0.196
PS Ventral Margin L	0.1633	0.33	0.1633	0.33
PS Ventral Margin R	0.3685	0.363	0.3685	0.363
PS Dorsal Margin L	0.9665	0.295	0.9665	0.295
PS Dorsal Margin R	0.1085	0.303	0.1085	0.303
SI Sup Topography L	0.0951	0.166	0.0951	0.166
SI Sup Topography R	0.2098	0.066	0.2098	0.066
SI Inf Topography L	<0.05*	0.128	0.7192	0.039
SI Inf Topography R	<0.05*	0.066	0.5375	0.016
SI Sup Morphology L	<0.05*	0.064	0.0969	0.117
SI Sup Morphology R	0.3823	0.01	0.3823	0.01
SI Apical Morphology L	0.1706	0.011	0.1706	0.011
SI Apical Morphology R	0.278	0.018	0.278	0.018
SI Inf Morphology L	<0.05*	0.104	0.1889	0.135
SI Inf Morphology R	<0.05*	0.072	0.068	0.078
SI Inf Texture L	0.1236	0.134	0.1236	0.134
SI Inf Texture R	0.0636	0.125	0.0636	0.125
SI Post Sup Exostoses L	0.6022	0.092	0.6022	0.092
SI Post Sup Exostoses R	0.672	0.088	0.672	0.088
SI Post Inf Exostoses L	<0.05*	0.007	NA	NA
SI Post Inf Exostoses R	<0.05*	0	NA	NA
SI Post Spicules L	<0.05*	0.004	0.8604	0.136
SI Post Spicules R	0.3229	0.001	0.3229	0.001
Coronal Pterica	0.7074	0.153	0.7074	0.153
Sagittal Obelica	<0.05*	0.068	0.2316	0.121
Lambdoidal Asterica	0.0765	0.077	0.0765	0.077
Interpalatine	<0.05*	0.188	0.7442	0.123
Zygomaticomaxillary	<0.05*	0.086	0.116	0.096

Table 22. Pooled male collapsed stages of traits for each age indicator

Age indicator	Traits	Before collapsing	After collapsing	Stages collapsed
Pubic symphysis	Relief L	5 stages	2 stages	Stage 2-4
	Relief R	5 stages	4 stages	Stage 3-4
	Texture L	4 stages	3 stages	Stage 2-3
	Texture R	4 stages	3 stages	Stage 1-2
Sacroiliac joint	Inferior topography L	3 stages	2 stages	Stage 2-3
	Inferior topography R	3 stages	2 stages	Stage 2-3
	Superior morphology L	4 stages	3 stages	Stage 3-4
	Inferior morphology L	4 stages	3 stages	Stage 3-4
	Inferior morphology R	4 stages	3 stages	Stage 3-4
	Posterior spicules L	3 stages	2 stages	Stage 1-2
Cranial sutures	Sagittal obelica	5 stages	2 stages	Stage 2-5
	Interpalatine	4 stages	3 stages	Stage 1-2
	Zygomaticomaxillary	5 stages	4 stages	Stage 4-5

Table 23. Pooled female Lagrange multiplier test and pseudo-R² before and after stage collapsing

Trait	Before stage collapsing		After stage collapsing	
	Goodness-of-fit	Pseudo-R ²	Goodness-of-fit	Pseudo-R ²
PS Relief L	<0.05*	0.08	0.2903	0.119
PS Relief R	<0.05*	0.107	0.2782	0.284
PS Texture L	0.3229	0.173	0.3229	0.173
PS Texture R	<0.05*	0.1	0.0631	0.201
PS Sup Apex L	<0.05*	0.273	0.8725	0.677
PS Sup Apex R	<0.05*	0.268	0.8725	0.677
PS Ventral Margin L	<0.05*	0.22	0.2349	0.248
PS Ventral Margin R	<0.05*	0.24	0.1865	0.231
PS Dorsal Margin L	<0.05*	0.136	0.6619	0.116
PS Dorsal Margin R	<0.05*	0.156	0.0646	0.22
SI Sup Topography L	<0.05*	0.124	0.5455	0.042
SI Sup Topography R	0.559	0.128	0.5674	0.177
SI Inf Topography L	0.6131	0.149	0.6131	0.149
SI Inf Topography R	0.842	0.132	0.842	0.132
SI Sup Morphology L	<0.05*	0.007	NA	NA
SI Sup Morphology R	<0.05*	0.018	NA	NA
SI Apical Morphology L	0.3959	0.049	0.3959	0.049
SI Apical Morphology R	0.2841	0.054	0.2841	0.054
SI Inf Morphology L	0.5403	0.169	0.5403	0.169
SI Inf Morphology R	0.9053	0.176	0.9053	0.176
SI Inf Texture L	<0.05*	0.218	0.0979	0.242
SI Inf Texture R	0.7104	0.129	0.7104	0.129
SI Post Sup Exostoses L	<0.05*	0.072	NA	NA
SI Post Sup Exostoses R	<0.05*	0.101	NA	NA
SI Post Inf Exostoses L	<0.05*	0.003	NA	NA
SI Post Inf Exostoses R	<0.05*	0.006	NA	NA
SI Post Spicules L	0.8871	0.006	0.8871	0.006
SI Post Spicules R	0.6604	0.014	0.6604	0.014
Coronal Pterica	0.1873	0.17	0.1873	0.17
Sagittal Obelica	0.0962	0.066	0.0962	0.066
Lambdoidal Asterica	0.5938	0.093	0.5938	0.093
Interpalatine	0.8329	0.198	0.8329	0.198
Zygomaticomaxillary	0.1305	0.038	0.1305	0.038

Table 24. Pooled females collapsed stages of traits for each age indicator

Age indicator	Traits	Before collapsing	After collapsing	Stages collapsed
Pubic symphysis	Relief L	5 stages	3 stages	Stage 1-2, 4-5
	Relief R	5 stages	4 stages	Stage 3-4
	Texture R	4 stages	3 stages	Stage 2-3
	Apex L	3 stages	2 stages	Stage 1-2
	Apex R	3 stages	2 stages	Stage 1-2
	Ventral margin L	7 stages	6 stages	Stage 5-6
	Ventral margin R	7 stages	6 stages	Stage 5-6
	Dorsal margin	5 stages	4 stages	Stage 1-2
	Dorsal margin	5 stages	4 stages	Stage 3-4
Sacroiliac joint	Superior topography L	3 stages	2 stages	Stage 2-3
	Inferior texture L	3 stages	2 stages	Stage 2-3

Table 25. RMSE for simulated males using TA scores (30 traits)

Age Indicator	Young (<40yrs) n=19	Middle (41-60yrs) n=92	Old (61yrs<) n=89
Pubic Symphysis (Uniform)	23.04	8.83	12.35
Sacroiliac Joint (Uniform)	21.52	10.32	11
Cranial Sutures (Uniform)	23.89	9.28	12.83
All (Uniform Prior)	13.8	10.28	10.51
All (Gompertz Prior)	14.5	15.16	9.69

Table 26. RMSE for simulated females using TA scores (27 traits)

Age Indicator	Young (<40yrs) n=11	Middle (41-60yrs) n=73	Old (61yrs<) n=116
Pubic Symphysis (Uniform)	19.55	8.12	13.59
Sacroiliac Joint (Uniform)	28.10	9.44	11.96
Cranial Sutures (Uniform)	23.66	8.95	13.53
All (Uniform Prior)	7.24	7.47	13.48
All (Gompertz Prior)	20.55	9.01	10.73

yielded most accurate age estimates. However, for middle-aged males, age estimates solely based on the pubic symphysis or the sacroiliac joint yielded the smallest RMSE value (9 years). Unlike the common belief, cranial sutures did not generate particularly poor age estimates. For simulated males, age estimates using a uniform prior yielded slightly better age estimates for the young (13 years) and the middle-aged (10 years) than those of an informed prior (14.5 years for the young, 15 years for the middle-aged). For old males, an informed prior seems to have minimal to no influence to improve age estimates.

For females, inaccuracy increased in order of the middle-aged, the old and the young. For young and middle-aged females, adding all three age indicators under a uniform prior yielded the most accurate age estimates (a RMSE value of 7 years in Table 26). Using the Gompertz prior rather increased inaccuracy in young and middle-aged females (20.5 years for the young, 9 years for the middle-aged). A role of an informed prior seemed to be more favorable for old females. Using a uniform prior, old females had a RMSE value of 13 years, while conditioning age estimates on a Gompertz hazard function reduced the RMSE value to 11 years.

Overall, pooled-sample models generated more accurate age estimates than sex- and population-specific models. Recall that there was a strong tendency of point estimates concentrated around middle-aged range. However, such tendency was alleviated in the pooled-sample models (Figures 34-37). Not only were the point estimates improved, but also estimated age ranges were shortened about 10-20 years (Table 27). In the previous analyses, the average widths of estimated age ranges were between 68-71 years for sex-population-specific samples. However, after the stage collapsing using pooled-male sample, the range narrowed down 10-20 years. Thus, despite the fact that it was a hypothetical informed prior, the Gompertz function contributed to narrow down age ranges for both Asian males and females: For females, an average

width of age estimates under a uniform prior is 65 years, and 60.5 years when the Gompertz function was used. For males, a uniform prior yielded age ranges with an average width 59 years, whereas incorporating a Gompertz function narrowed it down to 52 years.

In comparison to the age estimates of ADBOU, the pooled-Asian models generated improved age estimates for middle and old age groups than ADBOU. The most notable contribution of the pooled-Asian models is the reduced error for middle-aged adults (RMSE values were mostly under 10 years). For young adults, ADBOU performed better with about a 3-12 year smaller RMSE value than those of the pooled-Asian models. Thus, the pooled-Asian models still produced wide age ranges (average 65-52 years) compared to those of ADBOU (average 31-33 years). Nevertheless, it is important to note that although ADBOU generates narrower age estimates, it missed capturing a number of chronological ages within its age ranges (Figure 6, Figure 8), while the pooled-Asian models captured the known ages almost 100% (Figures 34-37).

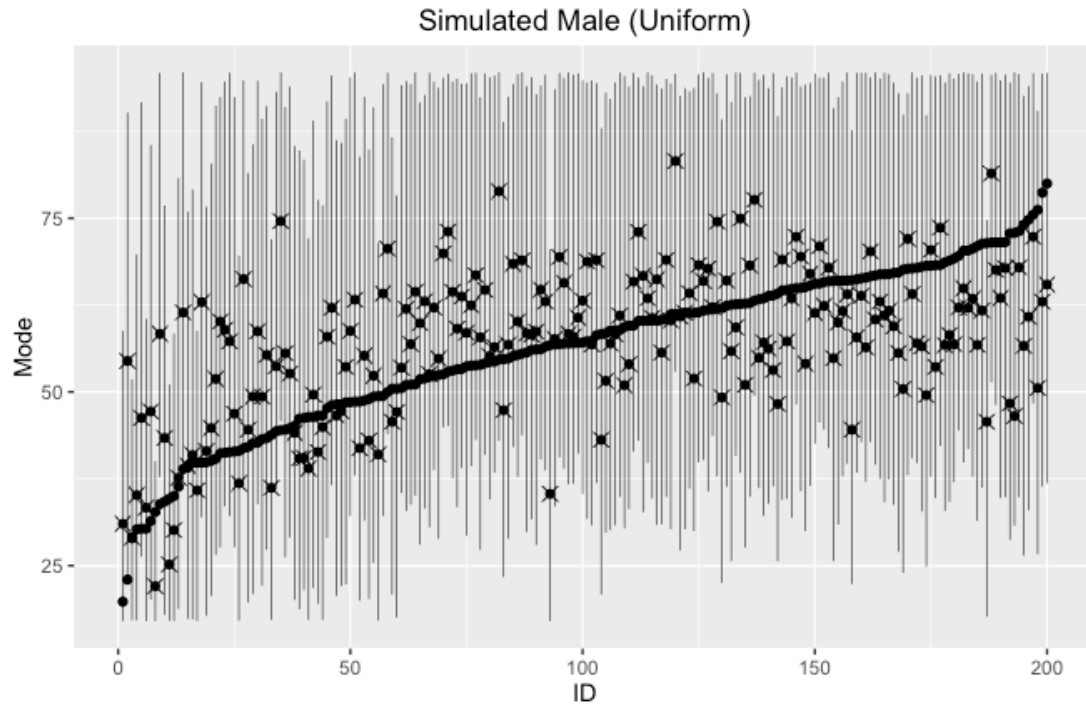


Figure 34. The pooled-male model age estimates under a uniform prior and documented chronological ages for simulated males (a uniform prior)

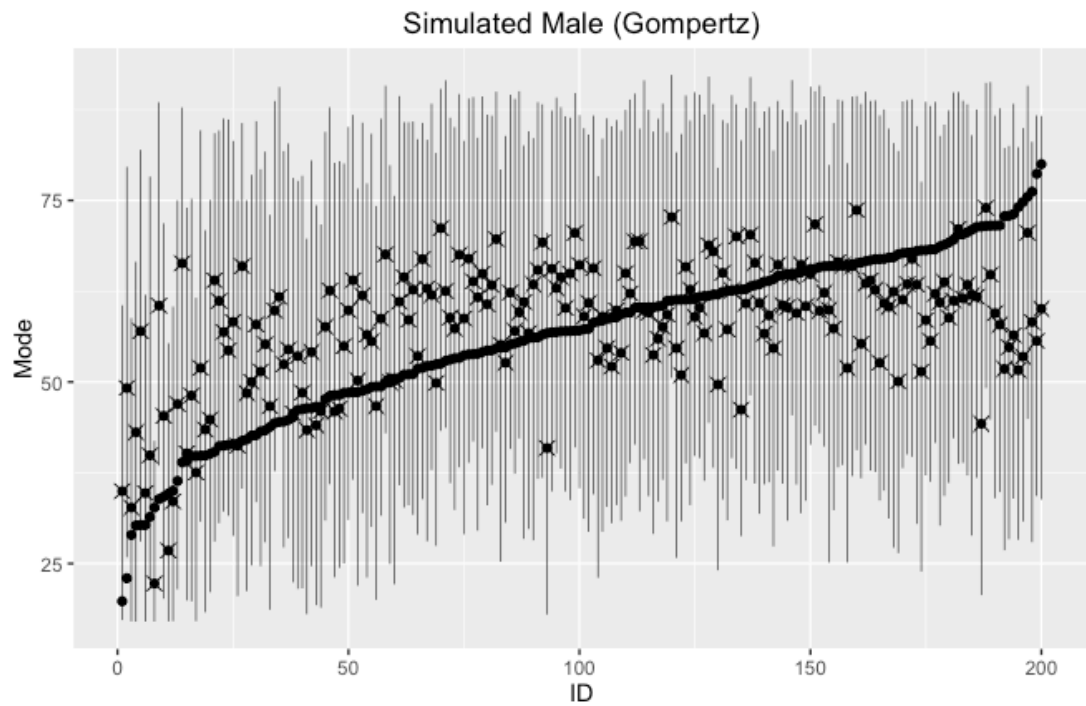


Figure 35. The pooled-male model age estimates and documented chronological ages for simulated males (an informed prior)

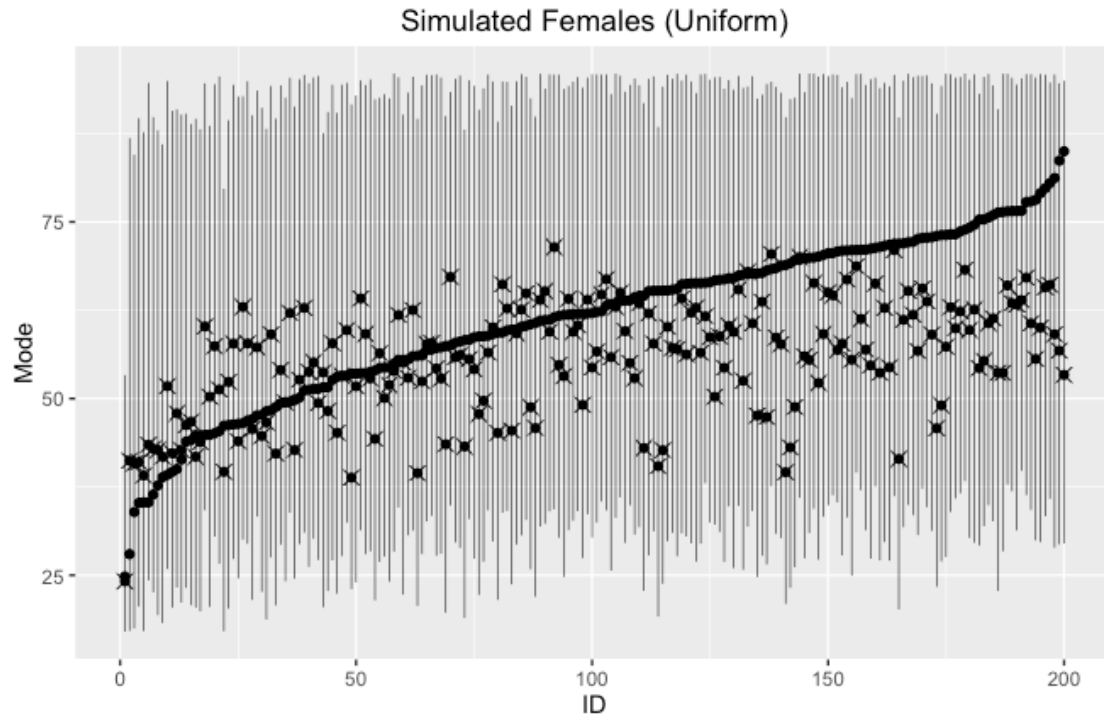


Figure 36. The pooled-female model age estimates and documented chronological ages for simulated females (a uniform prior)

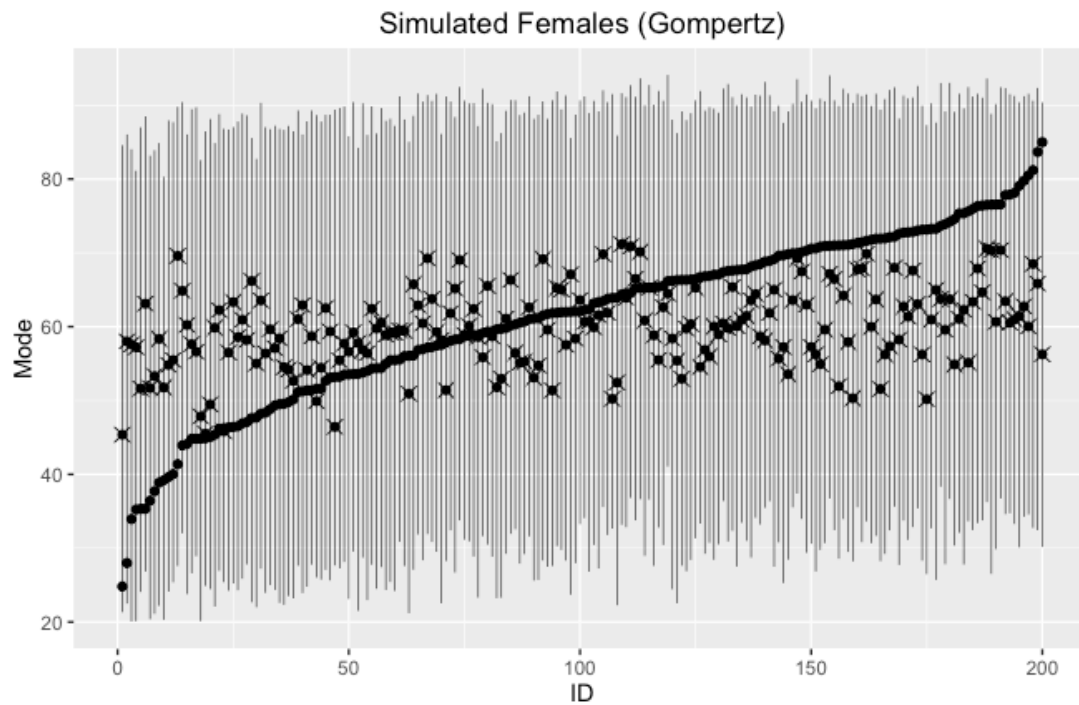


Figure 37. The pooled-female model age estimates and documented chronological ages for simulated females (an informed prior)

Table 27. Average widths of estimated age ranges for various models used in this study

	Uniform Prior	Gompertz Prior
ADBOU	32.94	31.22
Japanese Males	62.96	62.82
Thai Males	68.15	69.75
Japanese Females	67.69	67.09
Thai Females	71.15	71.13
Pooled Males	58.76	51.78
Pooled Females	64.75	60.50

Additional Test 1. Accuracy and reliability of conventional methods

In order to provide comparative information to the pooled-Asian models using TA scores, accuracy of three conventional aging methods on the Asian samples was evaluated. The conventional methods include Suchey-Brook's method of the pubic symphysis (Suchey and Katz 1998), Lovejoy et al.'s method of the sacroiliac joint (Lovejoy et al. 1985a), and Meindl and Lovejoy's method of cranial sutures (Meindl and Lovejoy 1985). To obtain final age estimates using the conventional methods' scores, the same analytical approach as Hypothesis 4 was applied. The multivariate probit regression model was fitted to the scores of the conventional methods (six variables including left and right sides) and relevant probit parameters were obtained. These parameters were then tested on a new target sample of 200 males and females that was simulated from a Gompertz function with $a_3=0.001$ and $b_3=0.01$.

Tables 28-29 show RMSE values for simulated males and females. Overall, age estimates incorporating all three conventional methods (or all three age indicators) produced the most accurate age estimates. Age estimates of all three methods combined for middle-aged males and females had the lowest RMSE values (6-7.5 years) among all other RMSE values observed so far. While the magnitude of error for young and old males was about the same (15 years), young

and old females had different RMSE values: Older females showed lower RMSE values (12-13 years) than the young females (15-17 years).

At the individual method level, Suchey-Brooks and Lovejoy et al's methods yielded the lowest error (18 years) for young males, and, for middle-aged males, it was Suchey-Brooks' method (5.5 years) followed by methods of Lovejoy et al. and Meindl and Lovejoy with the same amount of error (8 years). For old males, the three methods produced almost identical RMSE values (17-18 years). For females, the Lovejoy et al. method produced lowest error for the young (15 years) and old (7 years), while for the old females, both Suchey-Brooks and Meindl and Lovejoy's methods produced lowest RMSE values (15-15.5 years). Any effects of an informed prior in improving age estimates were not noticeable for both males and females.

In comparison to the pooled-Asian models built on TA scores, when age estimates of all three elements are considered, the conventional methods generated more accurate age estimates for the middle-aged adults, particularly for males, than TA (10-15 years for TA; 5.5-9 years for the conventional methods). However, TA performed better than the conventional methods for age estimation of the young (7-14 years for TA; 13-16 for the conventional methods), and the old (10-13 years for TA; 12-16 years for the conventional methods). In addition, estimated age ranges of conventional methods were very wide (not presented) as it was the case for the sex- and population-specific models.

Additional Test 2. Retrieving age estimates by combining TA and conventional methods

As the last part of the analysis, given that no scientific way of combining age estimates derived from multiple methods has been available, the analytical framework of the current study was

applied to scores of conventional methods and TA combined. This part evaluated if the multivariable ordered probit regression combined with an informed prior using the MCMC sampler can help alleviate the problem of obtaining one final set of age estimates when an observer is given with different estimates from multiple methods. Another valid reason to conduct this step is that, in the analysis of sex- and population-specific probit models, the results have shown that the greater number of traits used, the greater improvement in age estimates. Thus, carefully examining RMSE values of the pooled-Asian models built upon TA scores and those on conventional methods suggest that each of the models performs better for a particular age group. TA generated better age estimates for young adults and the old, while the conventional methods performed best for middle-aged adults. These results further indicate that age estimates can be potentially improved when TA and conventional methods are used in conjunction with each other.

To test this additional assumption, a multivariate probit regression model was fitted to scores of both TA and the three conventional methods from pooled-Asian males ($n=117$). Combining scores of TA and the conventional methods, 36 variables were fitted to the model. Because the size of the reference sample (117 Asian males) was small, not all stages were represented: The right ventral margin had no Stage 4, and the dorsal margin was missing Stage 2. Therefore, the two traits of TA were excluded from the analysis (see previous section on stage collapsing for the number of traits used for the pooled-Asian male model based on TA scores). Age estimation of pooled females was not conducted due to a small sample size ($n=42$) and a number of missing stages for the conventional methods. The probit parameters from the TA and conventional method scores of Asian males were conditioned upon a Gompertz function with

$a_3=0.001$ and $b_3=0.01$. Again, the TA-conventional model was tested on a simulated male sample ($n=200$).

The results indicated that when both TA and conventional methods were combined, there was surprisingly significant improvement in age estimates. RMSE values of the combined model are dropped below 10 years: 9 years for the young, 6 years for the middle-aged, and 9 years for the old (Table 30). One problematic issue with the sex- and population-specific models was that the models generate homogenous point estimates clustered around the middle-aged range. However, as shown in Figure 38, such a trend is absent in this combined model, and point estimates moved along the chronological ages. Figure 39 further reflects a decrease in error. Note that the strong U-shaped error distribution we saw in the sex- and population-specific models (Figures 28-29) is not evident in this figure. However, overestimation of the young and underestimation of the elderly persisted although the trend seemed to be alleviated.

There also was an improvement in estimated ranges. Recall that the average widths of age ranges for the pooled-male model using TA scores only were 52-59 years. However, when both TA and conventional methods' scores were combined, the average width of the age ranges reduced to 50 years. When compared to age ranges of ADBOU, a 50-year interval is still quite wide as ADBOU's average width of age ranges was 32-33 years. Nevertheless, it is worth noting that the pooled-male probit model with TA and the conventional methods combined captures 100% of the chronological ages (Figure 38), while ADBOU missed a number of chronological ages within its age range due to its shorter age interval (Figure 6, Figure 8). Additionally, when 50% highest posterior regions (HPDs) were estimated for pooled Asian males using the combined model, estimated age ranges became significantly narrower by 30 years than 95% HPDs: The 50% HPD region generated the average width of age ranges with 19 years. Figure 41

shows that the estimated age ranges are much shorter than Figure 38. However, now some of the intervals failed to capture known ages. The combined model with 50% HPDs missed chronological ages of 26 individuals out of 200, but, if interpreted differently, it still captures 87% of the documented ages within its range.

Table 28. RMSE for simulated males using conventional methods

	Young (<40yrs) n=13	Middle (41-60yrs) n=89	Old (61yrs<) n=98
Suchey-Brooks (Uniform)	17.86	5.54	17.10
Lovejoy et al. (Uniform)	17.76	7.85	17.69
Meindl & Lovejoy (Uniform)	20.51	7.64	17.72
All (Uniform Prior)	15.05	7.35	15.75
All (Gompertz Prior)	16.21	6.80	15.19

Table 29. RMSE for simulated females using conventional methods

	Young (<40yrs) n=11	Middle (41-60yrs) n=73	Old (61yrs<) n=116
Suchey-Brooks (Uniform)	23.15	9.39	15.21
Lovejoy et al. (Uniform)	15.02	6.94	16.98
Meindl & Lovejoy (Uniform)	18.47	8.54	15.56
All (Uniform Prior)	14.87	7.51	12.94
All (Gompertz Prior)	17.03	6.65	12.29

Table 30. TA and conventional methods combined: Pooled males parameters (n=117), simulated males (n=200)

	Young (n=19)	Middle (n=92)	Old (n=89)
RMSE	9.46	6.19	8.90
Ave. Age Range Width	Number of missed cases		
95% HPD	50.39	0 /200	
50% HPD	18.91	26 /200	

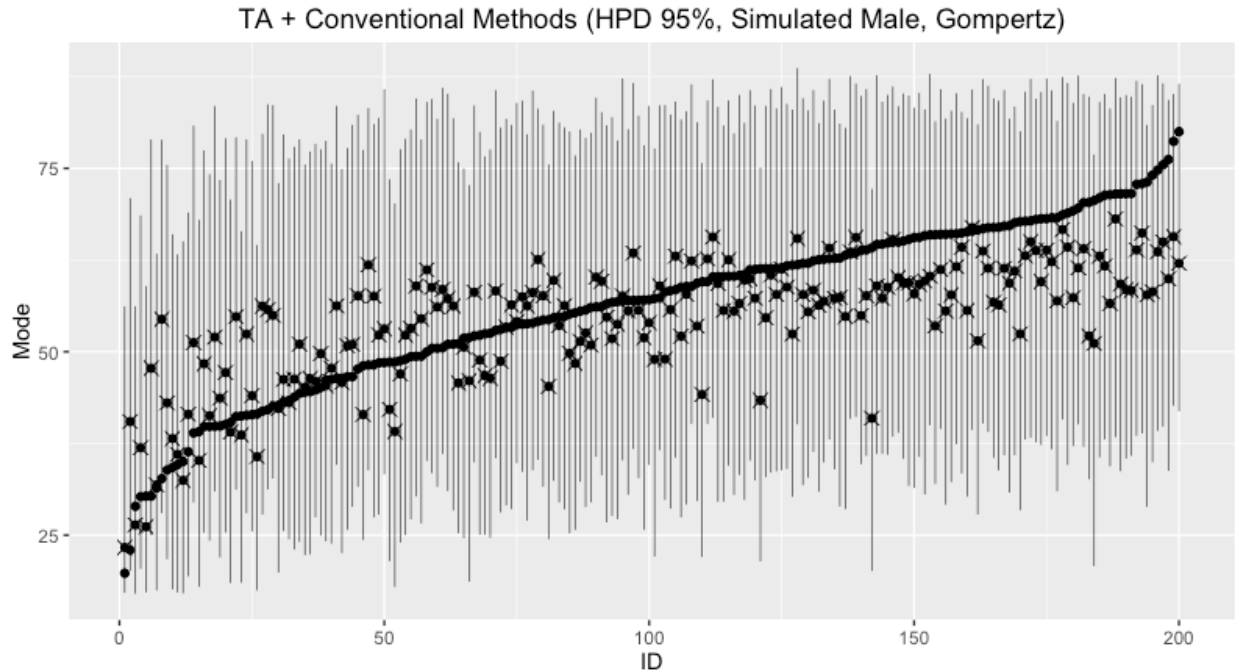


Figure 38. The TA-conventional method combined model age estimates and documented chronological ages for simulated males

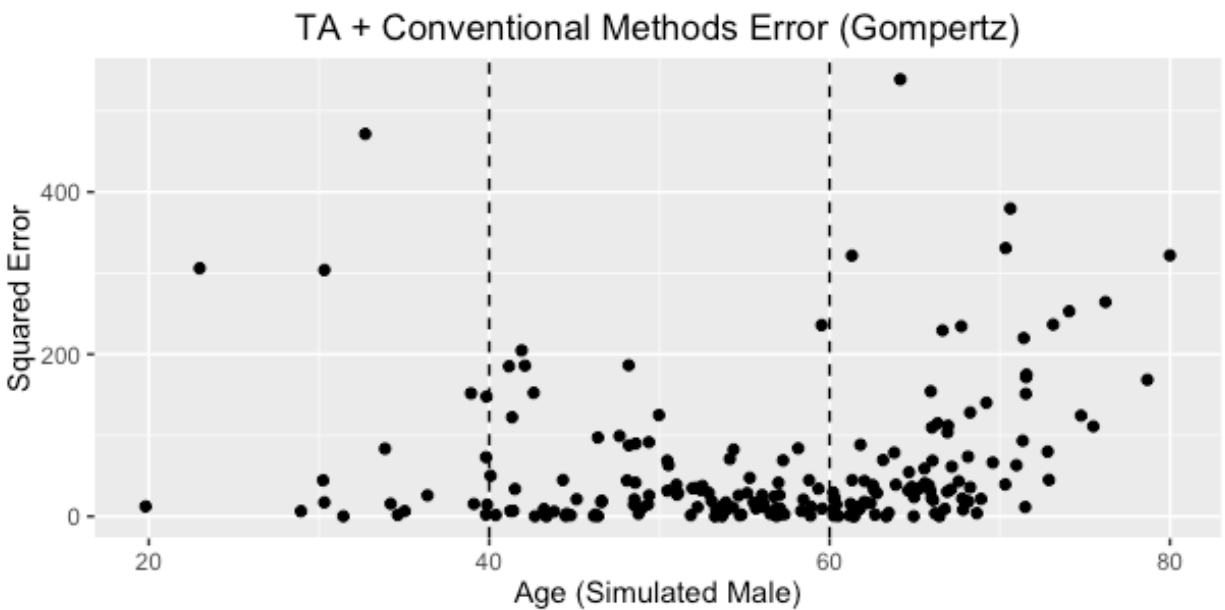


Figure 39. Error of the TA-conventional method combined model age estimates for simulated males

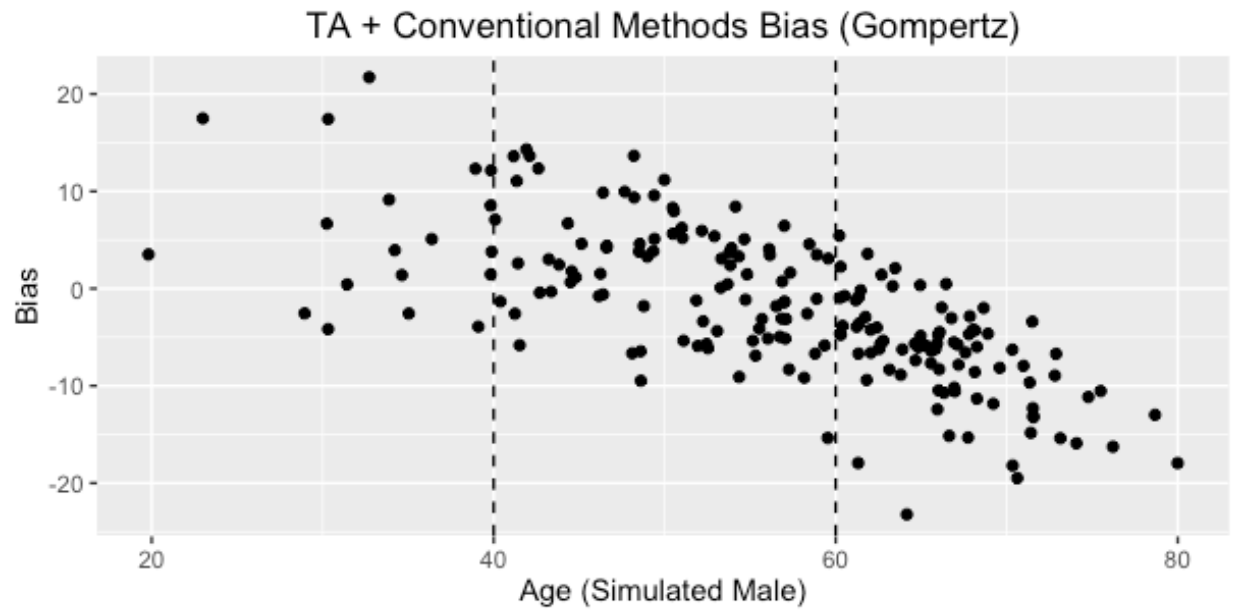


Figure 40. Bias of the TA-conventional method combined model age estimates for simulated males

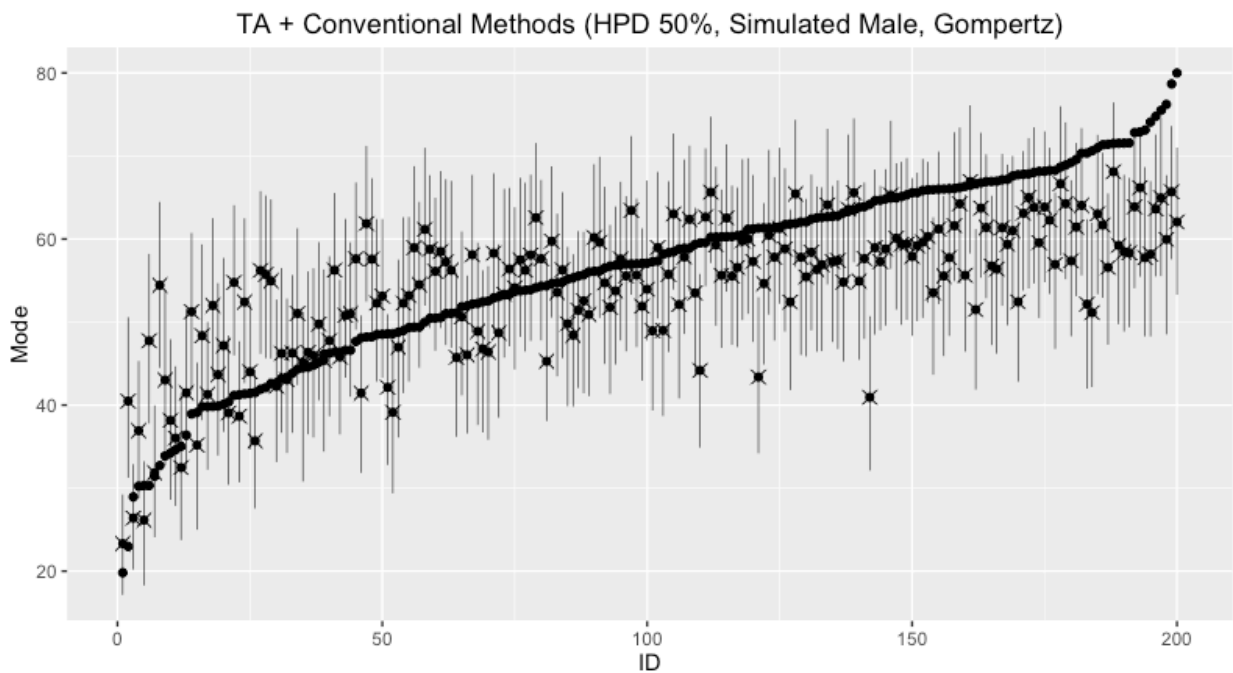


Figure 41. The TA-conventional method combined model age estimates and documented chronological ages for simulated males

Chapter 6

Discussion

The purpose of this study was to evaluate the utility of some current age estimation methods (TA and conventional methods) used for estimating ages at death within Asian populations, and to help understand population-specificity in skeletal aging rates and patterns of two Asian populations. To address the question of whether a single method can be developed by pooling two Asian populations (Japanese and Thais), four hypotheses were tested using four fully documented Asian skeletal samples based on the multivariate ordered probit regression under a Bayesian framework. The first hypothesis examined the performance of the TA age estimation program, ADBOU, on Asian populations. Hypotheses 2 and 3 investigated the effects of sex and population to age progressive changes in adult Asian skeletons. Hypothesis 4 attempted to help determine a breadth (or diversity) of populations that a single age estimation method can cover. In this chapter, the results presented in Chapter 5 are synthesized along with findings of previous studies. Each of the four hypotheses tested in the current study is also revisited. The first hypothesis concerning the ADBOU validation is discussed with the fourth hypothesis related to developing a pooled-Asian age estimation method.

Are there any sex differences in the schedule of skeletal aging?

Developmentally, female skeletons mature faster than male skeletons (Algee-Hewitt 2012; Kemkes-Grottenthaler 2002). Nevertheless, faster maturation does not necessarily result in faster degeneration, and this relationship is relatively under-explored. Moreover, while the current study has shown delayed aging rates of females, there is a lack of explanation in previous literature as to why females would age slower than males. A number of studies have attributed

faster degeneration of female pelvic joints to the reproductive function of females, pregnancy and childbirth. Gilbert and McKern (1973) noted that, as a fetus develops, the interpubic ligaments can be pulled extensively, causing hemorrhage, and leaving permanent pits or grooves (a.k.a. parity pits) on the dorsal surface of the pubis. In a more recent, extensive review of the pubic symphysis, Becker et al. (2010) state that during pregnancy, a circulating hormone called relaxin, in combination with estrogen, not only may increase the symphyseal width and mobility, but also may induce resorption of the symphyseal margins and structural changes in the fibrocartilaginous disc.

The second hypothesis investigated whether there are considerable differences between males and females in the schedule of skeletal aging among the Asian groups. The three symphyseal traits showed a mixed pattern in the schedule of trait transitions. For some traits both sexes experience the first transition about the same age, while for some others, females showed a faster transition from the first to second stage than males. After the first stage, females experience a delay in transition compared to males. Both left and right ventral margins and dorsal margins demonstrated slower aging in females than males after the first stage. However, there were two exceptions to note: The superior apex of Japanese males and females shared almost an identical schedule of transitions, whereas the same trait of Thais showed that Thai females showed a notably delayed, not faster, transition of the first stage than Thai males.

Before considering this trend in terms of biological reasons, it is important to restate the considerable sample size differences between males and females for this study. There are 138 Japanese males, but Japanese females are less than a half of the males ($n=49$). For the Thai sample, 157 Thai males and 75 Thai females were used for the analyses. Reexamining the sample distributions of Japanese and Thai males and females shows that not all age groups are

equally represented (Figure 42). Particularly, there are several age cohorts missing or underrepresented in the female samples. Both Japanese and Thai females are missing individuals in late 30s and the elderly (no late 80s in the Japanese female sample, and no early 90s in the Thai female sample). Especially, individuals under 40 years are significantly short in number for Thai females, in part explaining why the first mean age at transition in the superior apex of Thai females was estimated to be considerably later (age of 30 years) than males (age of 16-21 years).

While anthropologists agree that males and females undergo the same age-related sequential changes in skeletal morphology, no consensus has been reached whether or not there are sex differences in timing or tempo of skeletal aging. Todd (1921), the very first developer of the pubic symphysis aging method, found that there are negligible differences in aging between sexes. He noted that changes in the ventral aspect (ventral beveling, and ventral rampart completion) of females occur 2-3 years later than males. Conversely, the dorsal half flattens 2-3 years earlier than males, followed by an early commencement of dorsal margin lipping in females. In later years, Gilbert and McKern (1973) found the same trends in female pubic symphyses, but the amount of male-female differences was larger than what Todd had observed. The authors stated that the development of the ventral rampart is four years slower in females, whereas the dorsal half of the symphyseal face flattens faster up to 10 years than males. The current study, in part, echoes the two studies as there were delayed changes in the ventral half of the female symphysis. However, instead of faster dorsal margin flattening as the two previous studies demonstrated, both Japanese and Thai females exhibited a consistent delay in transition of the dorsal margin stages compared to males. This mixed trend in the schedule of pubic symphysis aging between Asian males and females could be either true innate sex differences or

sampling error, or both. However, for the current study, the latter scenario seems to be more plausible.

The functional stress to, and hormonal changes in, the pelvis have been presented as a main source of wider variability observed in female pelvic joints relative to those of males. And the wider variation in females is argued to increase inaccuracy in female age estimates (Brooks 1955; Hanihara and Suzuki 1978; Jackes 1985; Katz and Suchey 1989; Suchey and Katz 1998). Additionally, Berg (2008) and Hartnett (2010) noted that females are more susceptible to osteopenia and osteoporosis that may greatly contribute to accelerating aging rates. Yet, Hoppa (2000) counter-argues this statement. The author states that there are no significant differences in mean ages between low-birth vs. high-birth females, and therefore, observed differences between males and females are likely due to different age structures of study samples. Hoppa (2000) also found that female pubic symphyses above 40 years looked younger than males of the same age—an argument partly in line with the current study's results.

While studies of the pubic symphysis have detected some degree of sex differences in aging rates, other anatomical units, such as the sacroiliac joint and cranial sutures, did not render any significant difference between sexes (Buckberry and Chamberlain 2002; Lovejoy et al. 1985a; Meindl and Lovejoy 1985). Thus, echoing the initial findings of the original method by Lovejoy et al. (1985a), Igarashi et al. (2005) have demonstrated that rather than using sex specific equations, an equation developed on a pooled-sex sample yielded more accurate age estimates. Unfortunately, the current study could not assess sex differences on the rest of the age indicators, including the sacroiliac joint and cranial sutures, and the results of this study pose more questions than answers.

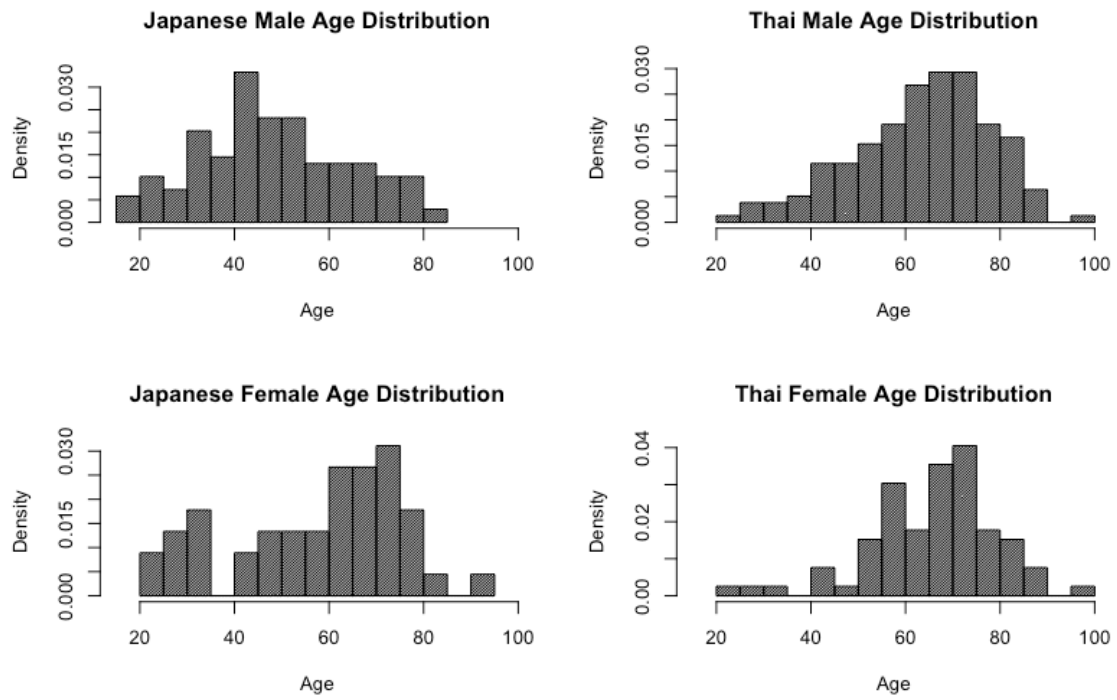


Figure 42. Age distributions of the current Asian sample

Are multiple population-specific methods an absolute necessity?

What was presented in Chapter 5, in part, challenges the assumption that each population needs an age estimation method exclusively specific to that group (Bullock et al. 2013; Djurić et al. 2007; Godde and Hens 2012; Katz and Suchey 1989; Komar 2003; Lottering et al. 2013; Schmitt et al. 2002). The current study has shown that the subcontinental population-specific models did not necessarily improve age estimates, and, at worse, yielded larger error for the target sample when age estimation was made based on too few variables (e.g. Thai male age estimates based on Thai male parameters had higher error rate than Thai male age estimates using Japanese male parameters). Instead, the pooled-Asian models generated significantly improved point and range estimates. A take-home point of this study is that the magnitude of error in age estimates is more associated with the number of age-informative traits used and the size and distribution of the reference sample rather than population variation.

One of the few proponents of the alternative view on population-specificity in skeletal aging is Lyle Konigsberg. Konigsberg et al. (2008) explicitly state, “[m]uch ink has needlessly been shed both in forensic and physical anthropology on the need for ‘population specific’ estimators...” (Konigsberg et al. 2008, p. 542). He and his colleagues then argue that what is seemingly between-group variation in aging of the skeleton is, in fact, artifacts of study samples with different (and often biased) age structures (Konigsberg et al. 2008, p. 542). Using a large dataset of pubic symphyseal scores from American, Balkan and Thai males (n= 1766), the authors showed that there are minimal between-population differences in ages at transition from one stage to the next higher one. They thus demonstrated that even if there is a slight difference in mean ages at transition between populations, it does not necessarily mean that one aging method developed on one population will perform poorly for other populations. For example,

there was a difference in mean ages at transition from Stage 2 to 3 between Balkan males (27.2 years) and non-Balkan males (21 years). To estimate ages of the Balkans, the two different parameters derived from the Balkans and non-Balkans were used. Despite the 6-year difference, there was a little impact on likelihood estimates or the posterior probabilities, and final age estimates were very similar regardless of using Balkan-specific or non-Balkan parameters. The current study has demonstrated strikingly parallel results to this argument. Although there were some differences in mean ages at transition between Thai and Japanese groups, final age estimates were very similar to each other.

In fact, arguing a lack of population differences in skeletal aging is not new. Hoppa (2000) investigated the performance of Suchey-Brooks' method on an American forensic sample and the 18th-19th century English cemetery population, and found that what seemingly looks like population-differences in skeletal aging is due to age structure differences between the reference and target samples, and/or secular change in modern vs. archaeological samples. Kimmerle et al. (2008) also did not find any skeletal aging differences between Balkan males and American males. Todd (1921) found that there are rather greater differences in skeletal aging between sexes than European descent vs. African descent. Thus, Murray and Murray (1991) claimed that inaccuracy of age estimates using Lovejoy et al.'s method is not dependent on ancestral or sex differences, but it is a more matter of using a single age indicator. Moreover, Hanihara and Suzuki (1978) specifically stated that the schedule of pubic symphysis aging is not different between Japanese and Americans. Lastly, TA, the method that lies at the center of this study, is built upon a reference sample made up of two geographically different populations, including Americans and Portuguese.

If it is not the effects of geographically distributed population, and if we eliminate the possibility of sampling error, then what causes (small or big) differences in skeletal aging between two distinct groups? One explanation is secular change. Although the study pertains to subadult and young adult age estimation, Langley-Shirley and Jantz (2010) present a comparative study of clavicle development using two American samples ranging from between 11 and 33 years of age, and representing 21st and 20th century populations. While the authors did not find any differences in skeletal aging between European Americans and African Americans, there was significant secular change between the two temporally distributed samples: The individuals from the 21st century showed faster fusion of the medial clavicle ends than the 20th century group. Shirley (2009) reviews an extensive volume of literature on factors that may influence the growth and development of the clavicles, thus contributing to secular change. According to Malina (1979), different rates of development observed in two different populations are due to differences in socioeconomic status intertwined with other growth-inhibiting factors, such as disease exposure, under-nutrition, and growth faltering. Schmeling et al. (2000) further argue that stature growth potential is a reflection of the child's health and nutrition combined with parents' cross-generational effects of nutritional status and morbidity. Based on these findings, Shirley (2009) concludes that the lack of difference in skeletal aging between young adults of African Americans and European American is possibly because the two ethnic groups shared the same environmental factors (e.g. socioeconomic status, nutrition, better public health, medical access, etc.) that govern growth and development.

Another explanation for between-group difference is population differences in physical activities and repetitive stress on the age indicators. The main idea has been that individuals with physically more demanding occupations or activity levels will show greater degeneration (i.e.

osteoarthritis) in major diarthroses with a greater degree of movement (Jurmain 2013). Although the pubic symphysis is a nonsynovial joint with a lack of movement, Cunningham et al. (2007) found frequent fibrocartilage herniation of the pubis among sports players, such as soccer players, who constantly apply severe biomechanical strain on the pubic symphysis and its supporting structures. This condition is associated with narrowing of the joint space in most cases and with para-symphyseal osteophytes in a minority of cases. However, Campanacho et al. (2012) found a little influence of physical activities to the pubic symphysis aging of Portuguese other than faster ligamentous outgrowths in the ventral aspect of the pubis.

More recently, obesity and its impact on degenerative changes in age indicators have been suggested. Wescott and Drew (2014) compared age estimates of a group of individuals of normal body mass index (BMI) with those with BMI greater than the normal level, using the auricular surface (Buckberry and Chamberlain 2002) and the pubic symphysis (Suchey-Brooks' method). Their study indicated that the obese group tends to be overestimated with increased inaccuracy than the normal BMI group. Merritt (2015) conducted a similar research using four different aging methods of the auricular surface, rib ends, pubic symphysis and five different measures of body size (i.e. BMI, stature, body mass, femoral length, femoral head diameter). The author concluded that short and light individuals tend to age slower, whereas tall and heavy individuals showed accelerated aging rates. Thus, the former group was consistently under-aged, while the latter was over-aged. Merritt (2015) attributes this trend to remodeling rates and bone mineral density, nutrition, physical activity and occupation, and hormone levels.

Another important point that the current study has shown is that both Japanese and Thais share almost identical characteristics of age-related traits. The quadrivariate probit model estimation indicated that Japanese and Thai individuals display analogous sets of traits from 20

to 90 years old. Age estimation is practiced under two major assumptions: invariance (Muller et al. 2002) and age monotonicity (Milner et al. 2008). Age monotonicity means that the sequential changes in skeletal characteristics are unidirectional and age progressive. The invariance means that the speed of skeletal aging stays constant throughout time and space—a concept also known as the uniformitarian assumption (Howell 1976)—and this second assumption has been at the basis of anthropologists applying a specific age estimation standard to another target population. The results of quadrivariate analysis have shown universal age-related skeletal characteristics, at least, across space.

Is there any improvement in age estimates of the new Asian models compared to ADBOU?

The current study has shown that there is significant improvement in age estimates derived from the pooled-Asian models in comparison to those produced by the current TA program, ADBOU. Particularly, error in age estimates of middle-aged adults, whose traits have been known to be notoriously difficult to characterize (Milner and Boldsen 2012), was greatly reduced to seven years compared to 13-14 years of ADBOU. Moreover, the probit model fitted to both TA and conventional scores combined produced even greater improvement in age estimates of the Asian samples. Another notable improvement other than reduced error was that the TA-conventional probit model corrected bias in age estimates about +/-30 years for the young and the old. Age estimates of ADBOU overestimated the young by about 50 years, while the old were underestimated by about 50 years. The TA-conventional model, on the other hand, showed 20-year overestimation of the young, and 20-year underestimation of the old. Additionally, a cautionary note for users of the current ADBOU is needed as the program consistently

underestimated Asian females, and, at worse, failed to capture chronological ages within its age range. If one wishes to build mortality profiles using age estimates of ADBOU, one should be aware of underestimation of female ages at death that might result higher young female mortality rates than one would expect.

Nevertheless, ADBOU performed better for young Asian adults, and produced narrower age ranges for all age groups compared to the pooled-Asian models or the TA-conventional combined model. Such unsatisfactory performance of the Asian probit model for the young mainly originated from the absence of young individuals below 40s in the current Asian sample. With such a biased sample, the models simply do not have enough information on young adults' characteristics, and consequently overestimation of the young will result. Thus, wide age ranges produced by the probit models of the Asian samples can be attributed to the fact that some skeletal traits are observed in a much broader age range than we thought. For example, Stage 3 and 4 of the left symphyseal relief are observed from 20s up to 80s in the Asian males (Figure 43). These traits add uncertainty to the age estimates, resulting in imprecise age estimates. This is not a new phenomenon as previous studies that have shown that there is much overlap between stages (Buckberry and Chamberlain 2002; Falys and Lewis 2011). Thus, the quadrivariate probit model demonstrated that, not only in a univariate scale, but also in a multivariate scale, age-related traits undergo a little to no changes from 40s up to 70s and 80s. As a result, elevated uncertainty in our age estimates of individuals within that age range will be evident. Milner and Boldsen (2012) have also pointed out that, during their 20-year long investigation to improve skeletal age estimation, they have not found any traits exclusively attributable to middle-aged adults. And that is why age estimates of middle-aged adults become very wide.

A wider age range is not always a negative aspect of aging methods. The Asian probit

models indeed produced wider age estimates than ADBOU. But the age estimates derived from the Asian probit models captured 100% of the chronological ages within its range, whereas ADBOU did not. Buckberry and Chamberlain (2002) recommend using wide age ranges over narrow ones, and state that wider estimates are more realistic as it allows to incorporate variations observed in trait expressions between individuals, and there is less chance to hinder identifications in forensic cases with incorrect age estimates. Moreover, another promising result of the current study is that when 50% highest posterior density region was obtained using the TA-conventional model, the age range narrowed down to, on average, 19 years, and still provided 87% accuracy.

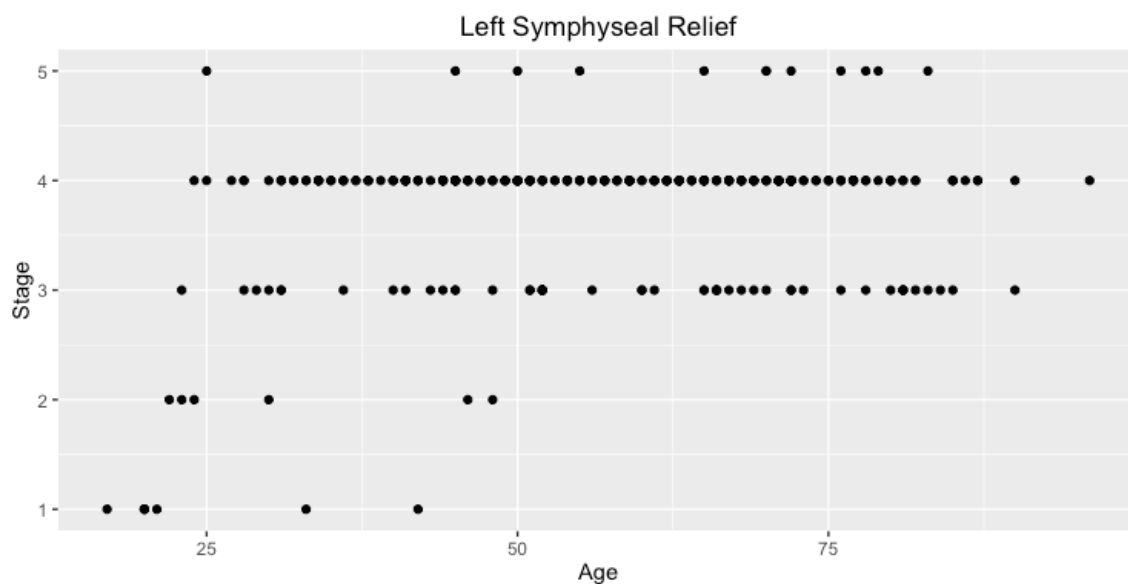


Figure 43. A score distribution of the left symphyseal relief showing Stages 3 and 4 cover a wide range of age cohorts

In addition, the study demonstrated that conventional methods do not necessarily yield higher error than TA. Previous findings have shown a poor performance of conventional methods when age estimates were retrieved using the methods' standard tables. However, the

results of this study showed that, when scores of conventional methods were analyzed based on the probit regression combined with Bayesian inferences, the methods provide useful information on ages at death of Asian populations.

More importantly, using all three age indicators improved age estimates, regardless of whether TA scores, conventional methods' scores, or both were used. A number of previous studies have confirmed that using multiple age indicators yields superior age estimates than using any single age indicator (Baccino et al. 1999; Martrille et al. 2007; Meindl et al. 1990; Saunders et al. 1992). What is new to note about the current study is that even when barely age informative traits, indicated by low pseudo R^2 values, are added to the model as an aggregate, they contributed to improving estimates.

The stage collapsing protocols of Konigsberg et al (2016) used in this study merits further discussion. The current study showed that dividing a trait into multiple stages does not always yield accurate and precise age estimates. For the Asian samples, 13 traits for males and 12 traits for females were subject to stage collapsing, and most of the stages were reduced to 2 stages. Although those stages 'had' to be collapsed in order to optimize the fit of the probit model, stage collapsing not only resulted in improving age estimates, but also increased correlations between skeletal traits and chronological ages. Such benefits of stage collapsing have been also supported in previous studies: Katz and Suchey (1986) collapsed Todd's 10-phase system of the pubic symphysis to a 6-phase system. Falys et al. (2006) reduced Buckberry and Chamberlain's 7-phase system, which is already a reduction of Lovejoy et al.'s 8-phase system for the sacroiliac joint, to 3 phases.

Recently, the developers of TA reconsidered their initial belief that the more number of stages, the better the age estimates. According to Milner and Boldsen (2012):

“When work on Transition Analysis was begun in 1996, we unthinkingly followed standard practice by focusing on the pubic symphysis, sacroiliac joint, and cranial sutures. The underlying assumption among osteologists, which we shared, was that useful skeletal structures had to display enough variation to allow the definition of multiple and sequential age-informative stages, the more the better. Each stage, of course, is associated with an age interval that overlaps with the next one. We are no longer convinced by the wisdom of that approach. In fact, we believe there is good evidence to suggest our initial focus on just cranial sutures and pelvic joints was far too narrow (p.11).”

The authors have shifted their focus to simple binary traits that are distributed throughout the skeleton and have low correlation with chronological ages. The authors claim that although each of the traits, when used independently, does not tell much about ages at death, when used as a group, it provides surprisingly more correct age estimates than ADBOU or current TA. They emphasize that the key to improving age estimation is to look beyond the three traditional age indicators. Although the current project is limited to the three traditional age markers, it supports Milner and Boldsen’s main argument that incorporating multiple traits, even if a trait’s relationship to chronological ages is weak, is a desirable way to pursue age estimation. Moreover, the stage collapsing procedure suggested by Konigsberg et al. (2015) provides a statistically robust way to reduce a number of stages when it is considered statistically necessary.

The current study has demonstrated that informed priors work in favor of narrowing down estimated age ranges, but did not do much to improve to point estimates. This finding is not limited to the current study, as any users of the current ADBOU might have noticed. ADBOU produces MLEs with very small differences before and after an informed hazard prior is selected. One might ask, then why we would use an informed prior. One undesirable outcome of using a uniform prior is that it can produce age estimates that are unrealistically old (Algee-Hewitt 2012). In ADBOU, it is not uncommon to see an age of 115 years as the upper end of an age range when age estimation was made based on a uniform prior. Conversely, computationally, when the size of an informed prior is bigger than the size of the data used for likelihood

estimates (i.e. reference Asian samples), the informed priors are known to override the posterior probabilities. However, in the current study, adding a Gompertz hazard model did not negatively overpower the point estimates, but rather contributed to narrowing the age range estimates.

Unless a justification can be made that the distribution of the target population is the same as the distribution of the reference sample, incorporating an informed prior as a reasonable guess to the target population's age structure is recommended.

Although the current study echoes many of the points made by Milner and Boldsen (2012) in their TA validation study, there are also a couple of non-parallel findings. First, Milner and Boldsen's study showed that the pubic symphysis performed best followed by the sacroiliac joint and cranial sutures. The authors argue that adding cranial sutures and the sacroiliac joint provides a little to no information on age estimation. At worse, the sacroiliac joint tends to underestimate older adults, while cranial sutures exacerbate inaccuracy. The validation of ADBOU on the Asian sample also showed that, when the sample was divided into sex- and population-specific groups, the individual age indicators performed best in order of the pubic symphysis, the sacroiliac joint, and cranial sutures. However, when the Asian sample was divided into three age cohorts, such a trend was absent. Instead, both the pubic symphysis and the sacroiliac joint performed equally well for young and middle-aged males, while the pubic symphysis and cranial sutures performed best for old males. For females, the sacroiliac joint produced the most accurate age estimates for young and middle-aged females, and the pubic symphysis for old females.

Error in age estimates of the pooled-Asian probit regression models showed that, for males, the sacroiliac joint fit best for the young and old, while all three age indicators worked well for the middle-aged. For females, the pubic symphysis yielded the lowest error for the

young, and the sacroiliac joint was most accurate for the old. Again, for middle-aged females, all three age indicators produced a similar amount of error. Lastly, all the three conventional methods worked well for the old males and females. For young males, the pubic symphysis and sacroiliac joint that produced the smallest error, while for middle-aged adults, the pubic symphysis worked best. For young and middle-aged females, the sacroiliac joint produced age estimates with the lowest error.

This list of age indicators' performance for each of the age groups should only be used as supplementary information, not as an absolute recommendation as to what indicator to use for age estimation of Asian skeletons. Either of these age indicators will yield *some* information on age, whether or not the method is phase-based or component-based. This study emphasizes, once again, that, whenever possible, multiple age indicators and multiple traits should be used for age estimation because multiple indicators/traits together will compensate different degrees of error between indicators/traits, and will eventually yield more accurate age estimates—although it is not uncommon to have only one skeletal element for age estimation during analysis of archeological and forensic skeletal remains. And the multivariate probit regression serves as a convenient tool to integrate information from multiple aging methods into synthesized final age estimates.

Limitations and future directions

Perhaps one of the biggest limitations of the study is a biased sample composition that does not represent all age groups. As a result, the Asian probit models could not adequately capture young individuals compared to ADBOU. An ideal reference sample should be close to a uniform distribution with everyone equally represented (Konigsberg et al. 2008). To improve the Asian

probit model, additional documented skeletal samples should fill in those age gaps. Thus, if possible, the factors that might have contributed to interfering natural biological skeletal aging, such as socioeconomic status and temporal differences, need to be controlled. Moreover, this study is limited to two Asian populations. Additional sampling of other Asian populations and non-Asian populations will further help clarify whether all Asians can be pooled into one group for a single method, and how much differences in aging rates are there between Asian groups and non-Asian groups.

A brief comment on the TA traits that are excluded from the pooled-Asian models is also necessary. Recall that three skeletal traits were excluded from the analysis of pooled-Asian males, and six traits were excluded from the analysis of pooled-Asian females. For males, the excluded traits included the left superior apex, and left and right posterior inferior exostoses. And, for females, it was the left and right superior morphology, and posterior superior and inferior exostoses. These traits were not included in the model because they were either not age informative (pseudo R^2 of zero), or the probit regression model could not be reasonably fitted (significant p values for the Lagrange multiplier test). The left superior apex of males seems to be the sampling error, and additional sampling can possibly fix the problem. This interpretation is more intuitive as the same trait for the opposite side and that of females worked well for modeling age progressive changes.

Nevertheless, the posterior exostoses of the sacroiliac joint pose a possibility that they are might not be a good age estimator after all for Asian populations. Buckberry and Chamberlain (2002) concluded that the posterior exostoses, the “retroauricular activity” in Lovejoy et al.’s term, were found to be a poor age estimator, and therefore eliminated from their revised method. Furthermore, Igarashi et al. (2005) found only bony bridging of the sacroiliac joint, that is

proposed to be an advanced stage of robust posterior exostoses, is more age informative than little scatters of bony outgrowths on the posterior aspect of the auricular surface. Nevertheless, sampling error cannot be confidently ruled out at this stage until more data are collected.

A third limitation is more technical. The current R functions to run a multivariate ordered probit regression cannot process missing values. Therefore, only can individuals with a complete suit of traits can be run for the analysis. This greatly reduced the sample size of the current study. The study started with 742 individuals, but after eliminating individuals with missing variables, the sample size boiled dipped to 419. While TA is designed to allow age estimation on incomplete elements, it is only possible when ADBOU is used, and the current multivariate probit regression scripts do not provide that option.

Finally, there are two recommendations to improve age estimation, not limited to this study, but for any studies on age estimation. The first recommendation involves objective trait evaluation to reduce observer error in age estimation. Whether or not TA or conventional methods, currently available age-at-death methods are based on visual assessment of age indicators. While this traditional approach is simple and straightforward once an observer familiarize oneself to age-related traits, such visual evaluation still introduces some degree of error within one observer and between observers. Boldsen et al. (2002b) attempted to reduce the issue of subjectivity by adopting a component scoring system and better quantifying trait descriptions, and the current study showed low intra-observer error, supporting high repeatability of the method. Thus, the most recent analysis on multiple rater reliability associated with TA scoring by Fojas et al. (2015) that involve five different observers has shown TA introduces minimal inter-observer error to age estimates. Lately, further attempts to reduce subjective judgments of observers in evaluating age-related traits started emerging. Slice and Algee -

Hewitt (2015) and Stoyanova et al. (2015) have introduced a series of fully computational age estimation methods by using 3D scanned images of the pubic symphysis and robust statistical algorithms that minimizes the effects of observers, and yet generates reliable age estimates for American males.

The second recommendation resonates the argument made by Milner and Boldsen (2012). To improve age estimates, we need to seek for a new set of skeletal traits throughout the entire skeleton. This study has shown that regardless of using advanced statistical models and combining multiple methods, there is limited information that the three traditional age indicators can offer to further enhance age estimates. This is not to say that these indicators need to be discarded in place of new indicators, as the three age indicators provide reasonable age-at-death estimates, and such explicit rejection undermines the foundational contributions of previous work. Instead, our next step is to look beyond traditional age markers, and to find novel age-related traits that are exclusive to a certain age cohort, so that when they are combined with the traditional age indicators, age estimates can be more precise and accurate.

In summary, the results of the current research, in part, supported Hypothesis 1. ADBOU produced point age estimates with greater error and bias than pooled-Asian probit models. However, estimated ranges of ADBOU were much narrower (but sometimes without capturing chronological ages) than pooled Asian probit models unless the highest posterior density region was reduced to 50% for age estimates derived from the pooled-Asian probit models. Hypothesis 2 was supported as there were some sex differences, but it is inconclusive, as the current study results cannot confidently rule out sample bias. Hypothesis 3 is fully supported as Japanese- and Thai-specific models did not necessarily yielded accurate age estimates for each group. Hypothesis 4 is also supported. Pooled-Asian models combining Japanese and Thais

significantly improved age estimates for Asian populations. Additionally, conventional methods did not yield particularly incorrect age estimates for Asian populations, and when the scores of conventional methods were combined with TA, the age estimates were the most accurate than any other models.

Chapter 7

Conclusion

The current research sought to quantify error and bias associated with currently available four age estimation methods for Asian populations, and to help understand population-specific age estimation using two geographically distributed Asian skeletal samples. This research ultimately attempted to evaluate how inclusive a single age estimation method can be in order to be applied to more diverse populations. As a first step to develop an inclusive age-at-death estimation for Asians, Japanese and Thai skeletal samples were pooled together to establish regression parameters that can be applied to both Japanese and Thai age estimation. The findings of this study can be summarized into four main points as listed below:

- 1) The pooled-Asian models performed the best over Thai- or Japanese-specific models, further suggesting that there is a lack of population-specificity in skeletal aging. Unlike the previous findings, an aging method developed on a specific population does not necessarily yield greater error for another population, and the error in age estimates is deemed not due to population differences. The increased error seems to be more associated with differences in age structures of the study sample, or a lack of age-informative traits used for age estimation.
- 2) More accurate age estimates occur when a larger number of traits are analyzed. Even though traits used for age estimation may have a weak correlation with chronological ages, when used in aggregate, they can produce accurate age estimates.
- 3) Although the effect of informed priors was not noticeable in improving point estimates, it is still recommended to use informed priors if possible as they help narrow estimated age ranges.
- 4) The multivariate ordered probit regression has an immediate application in combining multiple aging methods to yield synthesized final age estimates. Thus, despite the issue of subjectivity associated with conventional methods' phase-based approach, both phase-based and component-based methods equally yield reasonable age estimates for the Asian samples.

The results of this study have, once again, demonstrated that age estimation is meant to be erroneous because of the highly variable nature of human senescence (Bocquet-Appel and

Masset 1982; Schmitt et al. 2002), and because age at death estimation is only an approximation of chronological age using biological indicators as proxies (Milner et al. 2008). We have seen throughout this study's analyses that error induced from human variation is extremely difficult to eliminate even with an advanced statistical approach and combining multiple aging methods. We can only *reduce* such error with better practice, better research design and better statistics. Although skeletal traits are correlated with chronological ages to some degree, the correlations are not as high as one would hope. This speaks volumes that morphological changes of age indicators are not exclusively due to aging. Individually unique life histories leave individually idiosyncratic markers on age indicators, at the same time assigning different degrees of error for each skeleton (Boldsen et al. 2002b; Kemkes-Grottenthaler 2002; Milner et al. 2008). Although anthropologists have acknowledged there is great within-population variation as much as between-population variation, the focus has been placed on between-population variation. While there are numerous calls for developing population-specific age estimation methods under the assumption of population specificity in the skeletal aging, less emphasis has been placed on developing a more inclusive method that is broadly applicable to diverse (or unknown) populations. The current study has demonstrated that inaccuracy and bias we see in age estimates possibly do not stem from genetically and geographically predefined differences between groups. The most hopeful gesture of this study is that it demonstrated developing a single inclusive age estimation method, at least, for two Asian groups, Japanese and Thais, is possible.

Bibliography

- Acsádi G, and Nemeskéri J. 1970. History of human life span and mortality: Akadémiai Kiadó.
- Aeby C. 1858. Über die Symphyse ossium pubis des Menschen nebst Beiträgen zur Lehre vom hyalinen Knorpel und seiner Verknocherungen. *Z Rationelle Med Serie* 3(4):1-77.
- Aiello LC, and Molleson T. 1993. Are microscopic ageing techniques more accurate than macroscopic ageing techniques? *Journal of Archaeological Science* 20(6):689-704.
- Alesbury HS, Ubelaker DH, and 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 BFB. 2012. Age Estimation in Modern Forensic Anthropology. In: Tersigni-Tarrant MA, and Shirley NR, editors. *Forensic Anthropology: An Introduction*. Boca Raton, FL: CRC Press. p 181-230.
- Algee-Hewitt BFB, and Wilson-Taylor RJ. 2011. Getting Better With Age?: Testing the utility of Transition Analysis methods for forensic skeletal material of Hispanic origin. *The Annual Meeting of American Association of Physical Anthropologists*. Minneapolis, Minnesota.
- Anderson JY, and Trinkaus E. 1998. Patterns of sexual, bilateral and interpopulational variation in human femoral neck-shaft angles. *Journal of anatomy* 192(02):279-285.
- Aykroyd RG, Lucy D, Pollard AM, and Roberts CA. 1999. Nasty, brutish, but not necessarily short: a reconsideration of the statistical methods used to calculate age at death from adult human skeletal and dental age indicators. *American Antiquity* 64:55–70.
- Aykroyd RG, Lucy D, Pollard AM, and Solheim T. 1997. Technical note: Regression analysis in adult age estimation. *American Journal of Physical Anthropology* 104:259–265.
- Baccino E, Ubelaker DH, Hayek LC, and 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.
- Ballinger S, Schurr TG, Torroni A, Gan Y, Hodge J, Hassan K, Chen K, and Wallace DC. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130(1):139-152.
- Bartoo NW, and Parker KR. 1983. Stochastic age-frequency estimation using the von Bertalanffy growth equation. *Fishery Bulletin* 81(1):91-96.
- Baumhammers A, Stallard RE, and Zander HA. 1965. Remodeling of alveolar bone. *Journal of periodontology* 36(6):439-442.
- Becker I, Woodley SJ, and Stringer MD. 2010. The adult human pubic symphysis: a systematic review. *Journal of anatomy* 217(5):475-487.
- Bedford ME, Russell KF, Lovejoy CO, Meindl RS, Simpson SW, and Stuart-Macadam PL. 1993.

- Test of the multifactorial aging method using skeletons with known ages-at-death from the Grant collection. *American Journal of Physical Anthropology* 91:287–297.
- Bellwood P. 1996. Early agriculture and the dispersal of the southern Mongoloids. *Prehistoric Mongoloid Dispersals*. p 289-302.
- Berg GE. 2008. Pubic Bone Age Estimation in Adult Women. *Journal of forensic sciences* 53(3):569-577.
- Bethard JD. 2005. A Test of the Transition Analysis Method for Estimation of Age-at-Death in Adult Human Skeletal Remains.: University of Tennessee, Knoxville.
- Betti L, von Cramon-Taubadel N, Manica A, and Lycett SJ. 2013. Global Geometric Morphometric Analyses of the Human Pelvis Reveal Substantial Neutral Population History Effects, Even across Sexes. *Plos One* 8(2).
- Bocquet-Appel J, and Masset C. 1985. Paleodemography: resurrection or ghost? *Journal of Human Evolution* 14(2):107-111.
- Bocquet-Appel J-P. 1986. Once upon a time: Palaeodemography. *Mitteil Berlin Gesell Anthropol Ethnol Urges* 7:127-133.
- Bocquet-Appel J-P, and Masset C. 1982. Farewell to Paleodemography. . *Journal of Human Evolution* 11:321-333.
- Bogin B. 1999. Patterns of human growth. Cambridge, U.K. ; New York: Cambridge University Press. xiv, 455 p. p.
- Boldsen JL, Milner GR, and Hylleberg R. 2002a. ADBOU Age Estimator 2.0. ed. Odense, Denmark: The Anthropological Database Odense University, University of Southern Denmark.
- Boldsen JL, Milner GR, Konigsberg LW, and Wood JW. 2002b. Transition Analysis: A new method for estimating age from skeletons. In: Hoppa RD, and Vaupel JW, editors. *Paleodemography: Age distributions from skeletal samples*: Cambridge.
- Brace CL, and Hunt KD. 1990. A nonracial craniofacial perspective on human variation: A (ustralia) to Z (uni). *American Journal of Physical Anthropology* 82(3):341-360.
- Brace CL, Nelson AR, Seguchi N, Oe H, Sering L, Qifeng P, Yongyi L, and Tumen D. 2001. Old World Sources of the First New World Human Inhabitants: A Comparative Craniofacial View. *Proceedings of the National Academy of Sciences of the United States of America* 98(17):10017-10022.
- Brooks S, and Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human evolution* 5(3):227-238.
- Brooks ST. 1955. Skeletal age at death: the reliability of cranial and pubic age indicators. *Am J*

- Phys Anthropol 13(4):567-597.
- Buckberry JL, and Chamberlain AT. 2002. Age Estimation From the Auricular Surface of the Ilium: A Revised Method. *American Journal of Physical Anthropology* 119:231–239.
- Buikstra JE, and Konigsberg LW. 1985. Paleodemography: Critiques and Controversies. *American Anthropologist* 87:316-333.
- Buikstra JE, and Ubelaker DH. 1994. Standards for data collection from human skeletal remains.
- Bullock M, Márquez L, Hernández P, and 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*.
- Cameron J. 1982. Calibration. In: Katz S, and Johnson NL, editors. *Encyclopedia of Statistical Sciences*. New York: John Wiley and Sons. p 346-351.
- Campanacho V, Santos AL, and Cardoso HF. 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.
- Cavalli-Sforza LLL, Menozzi P, and Piazza A. 1994. *The history and geography of human genes*: Princeton university press.
- Clark W. 1981. Restricted least-squares estimates of age composition from length composition. *Canadian Journal of Fisheries and Aquatic Sciences* 38(3):297-307.
- Cleland J. 1889. On certain distinctions of form hitherto unnoticed in the human pelvis, characteristic of sex, age and race. *Memoirs and Memoranda in Anatomy* 1:95-103.
- Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–46.
- Committee on Identifying the Needs of the Forensic Sciences Community NRC. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, D.C. : the National Academies Press.
- Cunningham PM, Brennan D, O'Connell M, MacMahon P, O'Neill P, and Eustace S. 2007. Patterns of bone and soft-tissue injury at the symphysis pubis in soccer players: observations at MRI. *American Journal of Roentgenology* 188(3):W291-W296.
- de Aranzadi T. 1913. Algunas observaciones acerca del diagnóstico de la edad en el cráneo.
- Dempster AP, Laird NM, and Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society Series B (methodological)*:1-38.
- Dhanjal K, Bhardwaj M, and Liversidge H. 2006. Reproducibility of radiographic stage

- assessment of third molars. *Forensic science international* 159:S74-S77.
- Diamond J, and Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300(5619):597-603.
- DiGangi EA, Bethard JD, Kimmerle EH, and Konigsberg LW. 2009. A new method for estimating age-at-death from the first rib. *American Journal of Physical Anthropology* 138(2):164-176.
- Djurić M, Djonić D, Nikolić S, Popović D, and Marinković J. 2007. Evaluation of the Suchey–Brooks method for aging skeletons in the Balkans. *Journal of forensic sciences* 52(1):21-23.
- Dorandeu A, Coulibaly B, Piercecchi-Marti M-D, Bartoli C, Gaudart J, Baccino E, and Leonetti G. 2008. Age-at-death estimation based on the study of frontosphenoidal sutures. *Forensic science international* 177(1):47-51.
- Dudar JC, Pfeiffer S, and Saunders S. 1993. Evaluation of morphological and histological adult skeletal age-at-death estimation techniques using ribs. *Journal of forensic sciences* 38(3):677-685.
- Falys CG, and Lewis ME. 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 CG, Schutkowski H, and Weston DA. 2006. Auricular surface aging: worse than expected? A test of the revised method on a documented historic skeletal assemblage. *American Journal of Physical Anthropology* 130(4):508-513.
- Fojas CL, Kim J, Minsky-Rowland JD, and Algee-Hewitt B. 2015. Testing Inter-observer Reliability of the Transition Analysis Aging Method on the William M. Bass Forensic Skeletal Collection. 67th annual meeting of the American Academy of Forensic Sciences. Orlando, Florid. p 118.
- Fournier D, and Breen P. 1983. Estimation of abalone mortality rates with growth analysis. *Transactions of the American Fisheries Society* 112(3):403-411.
- Galton F. 1973. *Natural Inheritance* (1889). SERIES E: PHYSIOLOGICAL PSYCHOLOGY.
- Gamer M, Lemon J, Fellows I, and Singh P. 2012. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, and Hothorn T. 2014. Multivariate normal and t distributions. <http://cran.r-project.org/web/packages/mvtnorm/mvtnorm.pdf>.
- Gilbert BM, and McKern TW. 1973. A method for aging the female Os pubis. *American Journal of Physical Anthropology* 38(1):31-38.

- Glewwe P. 1997. A test of the normality assumption in ordered probit model. *Econometric Reviews* 16(1):1-19.
- Godde K, and Hens SM. 2012. Age-at-death estimation in an Italian historical sample: A test of the Suchey-Brooks and transition analysis methods. *Am J Phys Anthropol* 149(2):259-265.
- Gowland RL, and Chamberlain AT. 2002. A Bayesian Approach to Ageing Perinatal Skeletal Material from Archaeological Sites: Implications for the Evidence for Infanticide in Roman-Britain. *Journal of Archaeological Science* 29:677-685.
- Greene DL, Van Gerven DP, and Armelagos GJ. 1986. Life and death in ancient populations: bones of contention in paleodemography. *Human Evolution* 1(3):193-207.
- Hanihara K. 1991. Dual structure model for the population history of the Japanese. *Nichibunken Japan Review*:1-33.
- Hanihara K, and 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.
- Hanihara T. 1993. Craniofacial features of Southeast Asians and Jomonese: a reconsideration of their microevolution since the Late Pleistocene. *Anthropological science* 101(1):25-46.
- Hanihara T. 1994. Craniofacial continuity and discontinuity of Far Easterners in the Late Pleistocene and Holocene. *Journal of Human Evolution* 27(5):417-441.
- Hanihara T. 2006. Interpretation of craniofacial variation and diversification of East and Southeast Asians. In: Oxenham M, and Tayles NG, editors. *Bioarchaeology of Southeast Asia* Cambridge University Press, Cambridge. p 91-111.
- Hanihara T, and Ishida H. 2009. Regional Differences in Craniofacial Diversity and the Population History of Jomon Japan. *American Journal of Physical Anthropology* 139(3):311-322.
- Harris EF. 2007. Mineralization of the mandibular third molar: a study of American blacks and whites. *American journal of physical anthropology* 132(1):98-109.
- Hartnett KM. 2010. 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.
- Henle J. 1872. Bänder Zwischen beiden Hüftknochen. *Handbuch der Bänderlehre des Menschen (Handbuch der systematischen Anatomie)* 121.
- Higham C. 2001. Prehistory, language and human biology: is there a consensus in East and Southeast Asia. *Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia* World Scientific, Singapore:3-16.
- Hoppa RD. 2000. Population variation in osteological aging criteria: An example from the pubic

- symphysis. *American Journal of Physical Anthropology* 111(2):185-191.
- Hoppa RD, and Vaupel JW. 2002a. *Paleodemography : Age distribution from skeletal samples*. Cambridge ; New York: Cambridge University Press.
- Hoppa RD, and Vaupel JW. 2002b. The Rostock Manifesto. In: Hoppa RD, and Vaupel JW, editors. *Paleodemography: Age distributions from skeletal samples*: Cambridge.
- Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, and Pan I-H. 1996. mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *American journal of human genetics* 59(3):579.
- Howell N. 1976. Toward uniformitarian theory of human paelodemography. In: Ward RH, and Weiss KM, editors. *The Demographic Evolution of Human Populations*. New York, NY: Academic Press.
- Howells WW. 1986. *Physical anthropology of the prehistoric Japanese*. Windows on Japanese Past: Studies in Archaeology and Prehistory Center for Japanese Studies, The University of Michigan, Ann Arbor:85-99.
- Hunter W. 1762. Remarks on the symphysis of the ossa pubis. *Med Observations Inq* 2:333-339.
- Igarashi Y, Uesu K, T. W, and E. K. 2005. New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology* 128:324-339.
- İşcan. 1987. Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences* 32:452-466.
- İşcan M, Yoshino M, and Kato S. 1994. Sex determination from the tibia: standards for contemporary Japan. *Journal of forensic sciences* 39(3):785-792.
- İşcan MY, Loth SR, and Wright RK. 1984. Age estimation from the rib by phase analysis: white males. *Journal of Forensic Sciences* 29(4):1094-1104.
- İşcan MY, Loth SR, and Wright RK. 1985. Age estimation from the rib by phase analysis: white females. *Journal of Forensic Sciences* 30(3):853.
- İşcan MY, Yoshino M, and Kato S. 1995. Sexual dimorphism in modern Japanese crania. *American journal of human biology* 7(4):459-464.
- Ishida H. 1995. Nonmetric Cranial Variation of Northeast Asians and Their Population Affinities. *Anthropological Science* 103(4):385-401.
- Jacks M. 2000. Building the Bases for paleodemographic Analysis: Adult Age Determination. . In: Katzenber MA, and R. SS, editors. *Biological Anthropology of the Human Skeleton*. New York: John Wiley and sons, Inc. p 417-466.

- Jackes MK. 1985. Pubic symphysis age distributions. *Am J Phys Anthropol* 68(2):281-299.
- Japanese Mortality Database. 2014. National Institute of Population and Social Security Research. <http://www.ipss.go.jp/p-toukei/JMD/index-en.html> January 2015.
- Johansen H. 1998. Four early Danish parish registers. Research Report 6. Odense: Danish Center for Demographic Research.
- Johnson PA. 1996a. A test of the normality assumption in the ordered probit model. *Metron* 54:213-221.
- Johnson VE. 1996b. On Bayesian analysis of multirater ordinal data: An application to automated essay grading. *Journal of the American Statistical Association* 91(433):42-51.
- Jurmain R. 2013. *Stories from the skeleton: behavioral reconstruction in human osteology*: Routledge.
- Jurmain R, Kilgore L, Trevathan W, and Ciochon RL. 1994. *Introduction to physical anthropology*: West St Paul, MN.
- Katz D, and Suchey JM. 1986. Age determination of the male os pubis. *Am J Phys Anthropol* 69(4):427-435.
- Katz D, and Suchey JM. 1989. Race differences in pubic symphyseal aging patterns in the male. *Am J Phys Anthropol* 80(2):167-172.
- Kemkes-Grottenthaler A. 2002. Aging through the ages: Historical perspectives on age indicator methods. In: Hoppa RD, Vaupel, J. W. , editor. *Paleodemography: Age distributions from skeletal samples*. Cambridge.
- KemkesGrottenthaler A. 1996. Critical evaluation of osteomorphognostic methods to estimate adult age at death: A test of the "complex method". *Homo : internationale Zeitschrift fur die vergleichende Forschung am Menschen* 46(3):280-292.
- Khanpetch P, Prasitwattanseree S, Case DT, and Mahakkanukrauh P. 2012. Determination of sex from the metacarpals in a Thai population. *Forensic Science International* 217(1):229. e221-229. e228.
- Kimmerle EH, Konigsberg LW, Jantz RL, and Baraybar JP. 2008. Analysis of Age - at - Death Estimation Through the Use of Pubic Symphyseal Data. *Journal of forensic sciences* 53(3):558-568.
- Kimura DK. 1977. Statistical assessment of the age-length key. *Journal of the Fisheries Board of Canada* 34(3):317-324.
- Kimura DK, and Chikuni S. 1987. Mixtures of empirical distributions: an iterative application of the age-length key. *Biometrics*:23-35.

- King CA, Işcan M, and Loth SR. 1998. Metric and comparative analysis of sexual dimorphism in the Thai femur. *Journal of forensic sciences* 43(5):954-958.
- Komar D. 2003. Lessons from Srebrenica: the contributions and limitations of physical anthropology in identifying victims of war crimes. *Journal of Forensic Sciences* 48(4):713-716.
- Konigsberg L. 2015a. Lyle W. Konigsberg <https://50fc25581bb4561d048337efb1525036fe6ae84e.googleusercontent.com/host/0B-YNQGBRIWzxfmtwSUZreXpWTk1kVDA1TDVpVEZkd2ZhZnUzbEt6WjJDTTRINesyWU1ST0E/> 10 August 2015.
- Konigsberg L, and Frankenberg S. 1994. Paleodemography: "Not Quite Dead" *Evolutionary Anthropology* 3(92-105).
- Konigsberg L, and Herrmann N. 2002. Markov chain Monte Carlo estimation of hazard model parameters in paleodemography. In: Hoppa RD aVJ, editor. *Paleodemography: Age Distributions from Skeletal Samples*. New York, NY: Cambridge University Press. p 222-242.
- Konigsberg LW. 2015b. Multivariate cumulative probit for age estimation using ordinal categorical data. *Annals of human biology* 42(4):368-378.
- Konigsberg LW, and Frankenberg SR. 1992. Estimation of age structure in anthropological demography. *Am J Phys Anthropol* 89(2):235-256.
- Konigsberg LW, and Frankenberg SR. 2002. Deconstructing death in paleodemography. *American Journal of Physical Anthropology* 117(4):297-309.
- Konigsberg LW, and Frankenberg SR. 2013. Bayes in biological anthropology. *American journal of physical anthropology* 152(S57):153-184.
- Konigsberg LW, Frankenberg SR, and Liversidge HM. 2015. Optimal trait scoring for age estimation. *American journal of physical anthropology*.
- Konigsberg LW, Frankenberg SR, and Liversidge HM. 2016. Optimal trait scoring for age estimation. *American Journal of Physical Anthropology* 159(4):557-576.
- Konigsberg LW, Frankenberg SR, and Walker RB. 1997. Regress what on what? Paleodemographic age estimation as a calibration problem. *Integrating archaeological demography: multidisciplinary approaches to prehistoric population* Occasional Paper 24:64-88.
- Konigsberg LW, Hens SM, Jantz LM, and Jungers WL. 1998. Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. *Am J Phys Anthropol Suppl* 27:65-92.
- Konigsberg LW, Herrmann NP, Wescott DJ, and Kimmerle EH. 2008. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 53(3):541-557.

- Kruschke J. 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan: Academic Press.
- Landis JR, and Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics*:159-174.
- Langley-Shirley N, and Jantz RL. 2010. A Bayesian Approach to Age Estimation in Modern Americans from the Clavicle. . *Journal of Forensic Sciences* 55(571-583).
- Lanphear K. 1989. Testing the value of skeletal samples in demographic research: a comparison with vital registration samples. *International Journal of Anthropology* 4(3):185-193.
- Lewontin RC. 1972. The apportionment of human diversity. *Evolutionary biology*: Springer. p 381-398.
- Lewton KL. 2012. Complexity in Biological Anthropology in 2011: Species, Reproduction, and Sociality. *American Anthropologist* 114(2):196-202.
- Long JC. 2004. Human genetic variation: the mechanisms and results of microevolution. *Ann Arbor* 1001:48109.
- Long JC, and Kittles RA. 2009. Human genetic diversity and the nonexistence of biological races. *Human biology* 81(5/6):777-798.
- Lottering N, Macgregor DM, Meredith M, Alston CL, and Gregory LS. 2013. Evaluation of the suchey-brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. *Am J Phys Anthropol* 150(3):386-399.
- Lovejoy CO, Meindl RS, Pryzbeck TR, and Mensforth RP. 1985a. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology* 68(15-28).
- Lovejoy CO, Meindl RS, Pryzbeck TR, and Mensforth RP. 1985b. Multifactorial Determination of skeletal Age at Death: A Method and Blind Test of Its Accuracy. *American Journal of Physical Anthropology* 68:1-14.
- Lucy D, Aykroyd RG, Pollard AM, and Solheim T. 1996. A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. *Journal of Forensic Sciences* 41(2):189-194.
- Lucy D, and Pollard AM. 1995. Further comments on the estimation of error associated with the Gustafson dental age estimation method *Journal of Forensic Sciences* 40(2):222-227.
- Lynch SM. 2007. Introduction to applied Bayesian statistics and estimation for social scientists: Springer Science & Business Media.
- Mahakkanukrauh P, Khanpetch P, Prasitwattanseree S, and Case DT. 2013. Determination of sex from the proximal hand phalanges in a Thai population. *Forensic science international*.

- Mahakkanukrauh P, Khanpetch P, Prasitwattanseree S, Vichairat K, and Case DT. 2011. Stature estimation from long bone lengths in a Thai population. *Forensic Science International* 210(1-3).
- Malina RM. 1979. Secular changes in size and maturity: causes and effects. *Monographs of the Society for Research in Child Development*:59-102.
- Martin R. 1968. *Lehrbuch der Anthropologie—in systematischer Darstellung*.
- Martrille L, Ubelaker DH, Cattaneo C, Seguret F, Tremblay M, and Baccino E. 2007. Comparison of four skeletal methods for the estimation of age at death on white and black adults. *J Forensic Sci* 52(2):302-307.
- Masset C. 1989. Age estimation based on cranial sutures. In: Iscan MS, editor. *Age Markers in the Human Skeleton*. Springfield, IL: C.C. Thomas. p 71-103.
- Matsumura H. 2001. Differentials of Yayoi immigration to Japanas derived from dental metrics. *HOMO-Journal of Comparative Human Biology* 52(2):135-156.
- Matsumura H. 2006. The population history of Southeast Asia viewed from morphometric analyses of human skeletal and dental remains. In: Oxenham M, and Tayles NG, editors. *Bioarchaeology of Southeast Asia*. p 33-58.
- Matsumura H, and Hudson MJ. 2005. Dental perspectives on the population history of Southeast Asia. *American Journal of Physical Anthropology* 127(2):182-209.
- Matsumura H, and Oxenham M. 2013. Population Dispersal from East Asia into Southeast Asia: Evidence from Cranial and Dental Morphology. In: Pechenkina K, and Oxenham M, editors. *Bioarchaeology of East Asia*. Florida: University Press of Florida. p 179-209.
- Matsumura H, and Pookajorn S. 2005. A morphometric analysis of the Late Pleistocene human skeleton from the Moh Khiew Cave in Thailand. *HOMO-Journal of Comparative Human Biology* 56(2):93-118.
- Matsumura H, and Zuraina M. 1995. Metrical analysis of the dentition of Perak man from Gua Gunung Runtuh in Malaysia. *Bulletin of the National Science Museum, Tokyo, Series D* 21:1-10.
- Matsumura H, and Zuraina M. 1999. Metric analyses of an early Holocene human skeleton from Gua Gunung Runtuh, Malaysia. *American journal of physical anthropology* 109(3):327-340.
- McKern TW. 1957. Estimation of skeletal age from combined maturational activity. *American Journal of Physical Anthropology* 15(3):399-408.
- McKern TW, and Stewart TD. 1957. *Skeletal Age Changes in Young American Males: Analyzed from the Standpoint of Age Identification*. Technical Report EP-45. Natick, MA: US Army Quartermaster Research and Development Center.

- Meckel JF. 1820. Handbuch der menschlichen Anatomie: In den Buchhandlungen des Hallischen Waisenhauses.
- Meindl RS, and Lovejoy CO. 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 68:57-66.
- Meindl RS, Lovejoy CO, Mensforth RP, and Carlos LD. 1985a. Accuracy and direction of error in the sexing of the skeleton: implications for paleodemography. American Journal of Physical Anthropology 68(1):79-85.
- Meindl RS, Lovejoy CO, Mensforth RP, and Walker RA. 1985b. 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.
- Meindl RS, Russell KF, and Lovejoy CO. 1990. Reliability of age at death in the hamann - todd collection: Validity of subselection procedures used in blind tests of the summary age technique. American journal of physical anthropology 83(3):349-357.
- Mensforth RP. 1990. Paleodemography of the Carlston Annis (Bt - 5) Late Archaic skeletal population. American Journal of Physical Anthropology 82(1):81-99.
- Merritt CE. 2015. The influence of body size on adult skeletal age estimation methods. American journal of physical anthropology 156(1):35-57.
- Milner GR, and Boldsen JL. 2012. Transition analysis: A validation study with known-age modern American skeletons. American Journal of Physical Anthropology 148:98-110.
- Milner GR, and Boldsen JL. 2013. Transition Analysis Age Estimation: Skeletal Scoring Manual-Fordisc Version 1.00. University Park, PA USA, Odense, DK: The Pennsylvania State University, Syddansk Universitet.
- Milner GR, Wood JW, and Boldsen JL. 2008. Advances in Paleodemography. In: Katzenberg MA, and Saunders SR, editors. Biological Anthropology of Human Skeleton: John Wiley and Sons, Inc.
- Mukhopadhyay CC, Henze R, and Moses YT. 2013. How Real is Race?: A Sourcebook on Race, Culture, and Biology. Lanham, MD: Rowman & Littlefield.
- Muller HG, Love B, and Hoppa RD. 2002. Semiparametric method for estimating paleodemographic profiles from age indicator data. Am J Phys Anthropol 117(1):1-14.
- Murray KA, and Murray T. 1991. A Test of the Auricular Surface Aging Technique. . Journal of Forensic Sciences 36:1162-1169.
- Nagaoka T, and Hirata K. 2007. Reconstruction of paleodemographic characteristics from skeletal age at death distributions: perspectives from Hitotsubashi, Japan. American

- journal of physical anthropology 134(3):301-311.
- Nagaoka T, and 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.
- Nagaoka T, Hirata K, Yokota E, and Matsu'ura S. 2006. Paleodemography of a medieval population in Japan: Analysis of human skeletal remains from the Yuigahama - minami site. *American Journal of Physical Anthropology* 131(1):1-14.
- Nagaoka T, Sawada J, and Hirata K. 2008a. Did the Jomon people have a short lifespan? Evidence from the adult age-at-death estimation based on the auricular surface of the ilium. *Anthropological Science* 116(2):161-169.
- Nagaoka T, Shizushima A, Sawada J, Tomo S, Hoshino K, Sato H, and Hirata K. 2008b. Sex determination using mastoid process measurements: standards for Japanese human skeletons of the medieval and early modern periods. *Anthropological science* 116(2):105-113.
- Nei M. 1995. The origins of human populations: Genetic, linguistic, and archeological data. In: Brenner S, and Hanihara K, editors. *The origin and past of modern humans as viewed from DNA* World Scientific, Singapore. p 71-91.
- Olze A, Schmeling A, Taniguchi M, Maeda H, van Niekerk P, Wernecke K-D, and Geserick G. 2004. Forensic age estimation in living subjects: the ethnic factor in wisdom tooth mineralization. *International journal of legal medicine* 118(3):170-173.
- Omoto K. 1995. Genetic diversity and the origins of the "Mongoloids". In: Brenner S, and Hanihara K, editors. *The origin and past of modern humans as viewed from DNA*. Singapore: World Scientific. p 92-109.
- Omoto K, and Saitou N. 1997. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *American Journal of Physical Anthropology* 102(4):437-446.
- Osborne DL, Simmons TL, and Nawrocki SP. 2004. Reconsidering the auricular surface as an indicator of age at death. *Journal of Forensic Sciences* 49(5):905-911.
- Oxenham M, Matsumura H, Domett K, Thuy NK, Dung NK, Cuong NL, Huffer D, and Muller S. 2008. Health and the experience of childhood in Late Neolithic Viet Nam. *Asian Perspectives* 47(2):190-209.
- Oxenham M, and Tayles NG. 2006. *Bioarchaeology of Southeast Asia*. Cambridge, UK; New York: Cambridge University Press.
- Pechenkina K, Oxenham M, and Larsen CS. 2013. *Bioarchaeology of East Asia: movement, contact, health*. Florida: University of Florida Press.
- 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.
- Peters KD, Kochanek KD, and Murphy SL. 1998. Deaths: Final Data for 1996. National Vital Statistics Reports. Hyattsville, Maryland: National Center for Health Statistics.
- Pietrusewsky M. 1992. Japan, Asia and the Pacific: a multivariate craniometric investigation. In: Hanihara K, editor. *Japanese as a Member of the Asian and Pacific Populations*. International Research Center for Japanese Studies, Kyoto. Kyoto: International Research Center for Japanese Studies. p 9-52.
- Pietrusewsky M. 1994. Pacific-Asian relationships: a physical anthropological perspective. *Oceanic Linguistics* 33(2):407-429.
- Pietrusewsky M. 1999. A multivariate craniometric study of the inhabitants of the Ryukyu Islands and comparisons with cranial series from Japan, Asia, and the Pacific. *Anthropological Science* 107(4):255-281.
- Pietrusewsky M. 2006. A multivariate craniometric study of the prehistoric and modern inhabitants of Southeast Asia, East Asia, and surrounding regions: a human kaleidoscope. *Bioarchaeology of Southeast Asia* Cambridge: Cambridge University Press p:59-90.
- Pietrusewsky M. 2010. A multivariate analysis of measurements recorded in early and more modern crania from East Asia and Southeast Asia. *Quaternary International* 211(1):42-54.
- Pietrusewsky M. 2013. Biological Connections across the Sea of Japan: A Multivariate Comparison of Ancient and More Modern Crania from Japan, China, Korea, and Southeast Asia. In: Pechenkina K, and Oxenham M, editors. *Bioarchaeology of East Asia*. Florida: University Press of Florida. p 144-178.
- Piontek J, and Weber A. 1990. Controversy on paleodemography. *International Journal of Anthropology* 5(1):71-84.
- Prince DA, Kimmerle EH, and Konigsberg LW. 2008. A Bayesian approach to estimate skeletal age-at-death utilizing dental wear. *J Forensic Sci* 53(3):588-593.
- Quinn GP, and Keough MJ. 2002. *Experimental design and data analysis for biologists*: Cambridge University Press.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, and Johnson PL. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053-1060.
- Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM-S, Ko Y-C, Jinam TA, and Phipps ME. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics* 89(4):516-528.
- Relethford D, John H, and Harpending HC. 1994. Craniometric variation, genetic theory, and

- modern human origins. *American Journal of Physical Anthropology* 95(3):249-270.
- Relethford JH. 2001. Global analysis of regional differences in craniometric diversity and population substructure. *Human biology*:629-636.
- Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology* 118(4):393-398.
- Relethford JH. 2004a. Boas and beyond: migration and craniometric variation. *American journal of human biology : the official journal of the Human Biology Council* 16(4):379-386.
- Relethford JH. 2004b. Global patterns of isolation by distance based on genetic and morphological data. *Human biology* 76(4):499-513.
- Relethford JH. 2009. Race and global patterns of phenotypic variation. *Am J Phys Anthropol* 139(1):16-22.
- Ríos L, Weisensee K, and Rissech C. 2008. Sacral fusion as an aid in age estimation. *Forensic science international* 180(2):111.e111–111.e117.
- Rissech C, Wilson J, Winburn AP, Turbon D, and Steadman D. 2012. A comparison of three established age estimation methods on an adult Spanish sample. *International journal of legal medicine* 126(1):145-155.
- Roseman CC. 2004. Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proceedings of the National Academy of Sciences of the United States of America* 101(35):12824-12829.
- Roseman CC. 2016. Random genetic drift, natural selection, and noise in human cranial evolution. *American journal of physical anthropology*.
- Sakaue K. 2004. Sexual determination of long bones in recent Japanese. *Anthropological science* 112(1):75-81.
- Sakaue K. 2006. Application of the Suchey-Brooks system of pubic age estimation to recent Japanese skeletal material. *Anthropological Science* 114(1):59-64.
- Samworth R, and Gowland R. 2007. Estimation of adult skeletal age - at - death: statistical assumptions and applications. *International Journal of Osteoarchaeology* 17(2):174-188.
- Sashin D. 1930. A critical analysis of the anatomy and the pathologic changes of the sacro-iliac joints. *The Journal of Bone & Joint Surgery* 12(4):891-910.
- Saunders SR, Fitzgerald C, Rogers T, Dudar C, and McKillop H. 1992. A test of several methods of skeletal age estimation using a documented archaeological sample. *Journal of Forensic Science* 2:97-118.
- Savell KRR, Auerbach BM, and Roseman CC. 2016. Constraint, natural selection, and the

- evolution of human body form. *Proceedings of the National Academy of Sciences*.
- Schmeling A, Reisinger W, Loreck D, Vendura K, Markus W, and Geserick G. 2000. Effects of ethnicity on skeletal maturation: consequences for forensic age estimations. *International journal of legal medicine* 113(5):253-258.
- Schmitt A. 2004. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. *International Journal of Osteoarchaeology* 14:1-6.
- Schmitt A, Murail P, Cunha E, and 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 Sciences* 47(6):1203-1209.
- Schunke GB. 1938. The anatomy and development of the sacro - iliac joint in man. *The Anatomical record* 72(3):313-331.
- Shackelford LL, Harris S, Ashley E, and Konigsberg LW. 2012. Estimating the distribution of probable age - at - death from dental remains of immature human fossils. *American journal of physical anthropology* 147(2):227-253.
- Shi H, Zhong H, Peng Y, Dong Y-L, Qi X-B, Zhang F, Liu L-F, Tan S-J, Ma RZ, Xiao C-J et al. . 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biology* 6(1):1-10.
- Shirley NR. 2009. Age and sex estimation from the human clavicle: An investigation of traditional and novel methods [Doctoral Dissertation]: University of Tennessee, Knoxville.
- Shirley NR, and Ramirez Montes PA. 2015. Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component - Based Methods. *Journal of forensic sciences* 60(1):107-111.
- Singer R. 1953. Estimation of age from cranial suture closure. A report on its unreliability. *Journal of Forensic Medicine* 1(1):52-59.
- Sinha A, and Gupta V. 1995. A study on estimation of age from pubic symphysis. *Forensic science international* 75(1):73-78.
- Slice DE, and Algee - Hewitt BF. 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.
- Smith HF. 2009. Which cranial regions reflect molecular distances reliably in humans? Evidence from three - dimensional morphology. *American Journal of Human Biology* 21(1):36-47.
- Steadman DW, Adams B, and Konigsberg L. 2006. The Statistical Basis for Positive

- Identifications In Forensic Anthropology. *American Journal of Physical Anthropology* 131(1):15-26.
- Stinson S. 2012. Growth variation: biological and cultural factors. In: Stinson S, Bogin B, and Dennis OR, editors. *Human biology: an evolutionary and biocultural perspective* New York: Wiley-Liss p. p 587-635.
- Stout S. 1986. The use of bone histomorphometry in skeletal identification: the case of Francisco Pizarro. *Journal of forensic sciences* 31(1):296.
- Stout SD. 1992. Methods of determining age at death using bone microstructure. *Skeletal biology of past peoples: research methods* New York: Wiley-Liss:21-35.
- Stout SD, and Paine RR. 1992. Brief communication: Histological age estimation using rib and clavicle. *American Journal of Physical Anthropology* 87(1):111-115.
- Stoyanova D, Algee - Hewitt BF, and Slice DE. 2015. An enhanced computational method for age - at - death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *American journal of physical anthropology* 158(3):431-440.
- Suchey JM, Brooks ST, and Katz D. 1986. Instructions for use of the Suchey–Brooks system for age determination of the female os pubis. Instruction materials accompanying female pubic symphyseal models of the Suchey-Brooks system. Fort Collins, CO: France Casting.
- Suchey JM, and Katz D. 1998. Applications of pubic age determination in a forensic setting. *Forensic osteology: advances in the identification of human remains* Springfield, IL: Charles C Thomas 2.
- SWGANTH. 2013. Scientific Working Group for Forensic Anthropology: Age Estimation p1-5.
- Tan S. 2001. Genetic relationships among sixteen ethnic groups from Malaysia and Southeast Asia. In: Jin L, Seielstad M, and Xiao C, editors. *Genetic, Linguistic and Archeological Perspectives on Human Diversity in Southeast Asia* World Scientific, Singapore. p 83-91.
- Tayles NG, and Oxenham M. 2006. Introduction: Southeast Asian bioarchaeology past and present. In: Oxenham M, and Tayles NG, editors. *Bioarchaeology of Southeast Asia*: Cambridge University Press.
- TenCate A, and Mills C. 1972. The development of the periodontium: the origin of alveolar bone. *The Anatomical record* 173(1):69-77.
- Todd TW. 1920. Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology* 3(3):285-334.
- Todd TW. 1921. Age changes in the pubic bone. *American Journal of Physical Anthropology* 4(1):1-70.

- Todd TW, and Lyon D. 1924a. 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 TW, and 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 TW, and Lyon D. 1925b. Suture closure—its progress and age relationship. Part IV.—Ectocranial closure in adult males of Negro stock. *American Journal of Physical Anthropology* 8(2):149-168.
- Todd TW, and Lyon DW. 1924b. Endocranial suture closure. Its progress and age relationship. Part I.—Adult males of white stock. *American Journal of Physical Anthropology* 7(3):325-384.
- Turner C. 1989. Teeth and prehistory in Asia. *Scientific American* 260(2):88-91, 94-86.
- Turner C. 1990. Major features of Sundadonty and Sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian aboriginals. *American Journal of Physical Anthropology* 82(3):295-317.
- Turner C. 1992a. Microevolution of East Asian and European populations: a dental perspective. In: Aoki K, and Kimura T, editors. *The Evolution and Dispersal of Modern Humans in Asia*. Tokyo: Hokusensha. p 415-438.
- Turner C. 1992b. Sundadonty and Sinodonty in Japan: the dental basis for a dual origin hypothesis for the peopling of the Japanese Islands. In: Hanihara K, editor. *Japanese as Member of the Asian and Pacific Populations, International Symposium*. p 96-112.
- U.S. Census Bureau-Population Division. 2013. Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States, States, and Counties: April 1, 2010 to July 1, 2012. Release Date: June 2013.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, and Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics* 65(1):43-62.
- Usher BM. 2002. Reference Samples: the first step in linking biology and age in the human skeleton. In: Hoppa RD, and Vaupel JW, editors. *Paleodemography: Age distributions from skeletal samples*. Cambridge.
- Van Gerven DP, and Armelagos GJ. 1983. “Farewell to paleodemography?” Rumors of its death have been greatly exaggerated. *Journal of Human evolution* 12(4):353-360.
- Vaughn B, and Morrison WM. 2006. *China-Southeast Asia Relations: Trends, Issues, and Implications for the United States*. DTIC Document.

- von Cramon-Taubadel N. 2014. Evolutionary insights into global patterns of human cranial diversity: population history, climatic and dietary effects. *J Anthropol Sci* 92:43-77.
- Weiss AA. 1997. Specification tests in ordered logit and probit models. *Econometric Reviews* 16(4):361-391.
- Welcker N. 1866. *Kraniologische Mitteilungen (Archiv f. Anthropologie, I)*. Braunschweig DE JM JUDT DIE JUDEN ALS EASSE 235.
- Wescott DJ, and Drew JL. 2014. Effect of obesity on the reliability of age - at - death indicators of the pelvis. *American journal of physical anthropology*.
- Wescott DJ, and Drew JL. 2015. Effect of obesity on the reliability of age - at - death indicators of the pelvis. *American journal of physical anthropology* 156(4):595-605.
- Wilson RJ, and Algee-Hewitt BFB. 2009. [Inter] facing age: A test of the ADBOU age estimation software in a forensic context. *The Annual Meeting of the American Academy of Physical Anthropologists*. Chicago, IL.
- Yamaguchi B. 1992. Skeletal morphology of the Jomon people. *Japanese as a Member of the Asian and Pacific Populations, International Symposium*. p 52-63.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, and Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70.

Vita

Ji Eun Kim was born and raised in Seoul, South Korea until 18 years of age. Being apart from her family, her long journey to the U.S. started on September 2nd, 2002 at the JFK airport, New York, NY, with her little suitcase. After completing a series of English as second language programs at New York University and City University of New York-Baruch College, she was accepted to State University of New York at Binghamton in 2004. When she was a sophomore at SUNY Binghamton, she met Dr. Dawnie Steadman like a destiny through a course, named 'Bones, Bugs and Forensic Sciences,' and was converted to an osteology enthusiast, and has been part of the 'Wolfe' pack ever since. She completed her BA degree in 2008 and an MA degree in 2011 at SUNY Binghamton. Ji Eun Kim moved to Knoxville, TN, in 2011 to pursue a doctoral degree at University of Tennessee, Knoxville, under the continued guidance of Dr. Dawnie Steadman. She will complete her Ph.D. in December 2016. She will start postdoctoral research from September 2016 at Florida State University, and is looking forward to another new journey to unfold.