2009

An Examination of Error in the Application of Pubic Aging Techniques

Ellery Kate Lungmus
The University of Montana

AN EXAMINATION OF ERROR IN THE APPLICATION OF PUBIC AGING TECHNIQUES

By
Ellery Kate Lungmus
B.A. Anthropology, St. Lawrence University, New York, 2008

Thesis presented in partial fulfillment of the requirements for the degree of

Master of Arts in Anthropology, Forensic Anthropology

The University of Montana
Missoula, MT

December 2009

Approved by:

Perry Brown, Associate Provost for Graduate Education
Graduate School

Ashley McKeown, Chair
Anthropology

Randall Skelton
Anthropology

Christiane von Reichert
Geography

Lungmus, Ellery, M.A., Fall 2009
Anthropology

An Examination of Error in the Application of Pubic Aging Techniques

Chair: Ashley McKeown

This study examined six methods of skeletal age-at-death estimation from the pubic symphysis in order to determine the significance of sex as a contributing source of error to the inaccuracy of each method. These six methods included Todd (1920), McKern-Stewart (1957), Gilbert-McKern (1973), Hanihara-Suzuki (1978), Suchey-Brooks female-specific (Brooks and Suchey 1990), Suchey-Brooks male-specific (Brooks and Suchey 1990), and the Berg female-specific Suchey-Brooks 7^{th} phase addition (2008).

Three hundred and ninety-six individuals were randomly selected from the William M. Bass (WMB) Donated Skeletal Collection housed at The University of Tennessee, Knoxville, and were evaluated without knowledge of actual age for age-at-death in six observations, one for each method. These data were combined, re-associated with data on the age and sex of each individual, and both bias and inaccuracy were calculated for each method. Independent samples t-tests for equality of means were used to determine the significance of the difference between mean bias and inaccuracy across male and female sex categories.

The results of this study suggest that average inaccuracy was not significantly different between males and females for any of the tested methods. This is interesting considering the extensive body of research that has suggested that the range of variation for the female pubic symphysis is greater than for males for reasons such as dimorphic pelvic morphology, parturition (childbirth), and greater rates of osteoporosis. The Berg 7^{th} phase addition to the Suchey-Brooks method did not perform as well as initial tests suggested, and based on the results of this study, this addition should not be considered preferable to the original six-phase Suchey-Brooks female-specific method. Overall, the two Suchey-Brooks methods performed best in terms of coverage, though the results of this study suggest that male- and female-specific methods may not significantly improve accuracy.

As aging techniques based on isolated American samples are increasingly globally applied, research on the specific nature of the weaknesses of each method becomes critical. It is hoped that the results of this study will help to clarify the contribution of sex as a proposed source of error in addition to providing direction for further research.

ACKNOWLEDGMENTS

The completion of this program and specifically this project is due to the assistance and support of many contributors. The author wishes to thank the Office of the Vice President for Student Affairs at The University of Montana and the Dean of Students for friendship and support; the Office of the President of The University of Montana and the Excellence Fund for travel and research assistance; the William M. Bass Donated Collection at The University of Tennessee, Knoxville and its staff, including Dr. Lee Meadows Jantz, Heli Maijanen, and Rachel Wilson for access to the collection and research support; Dr. Ashley McKeown for supporting the development of this project and Drs. Randall Skelton and Christiane von Reichert for seeing it through to completion; the Maureen and Mike Mansfield and Tozzer Libraries for research support; and Brian Ouellette for statistical crisis triage and his infinite patience.

This project is dedicated to Laura and Edgar Job, who would have been tickled to see it happen.

TABLE OF CONTENTS

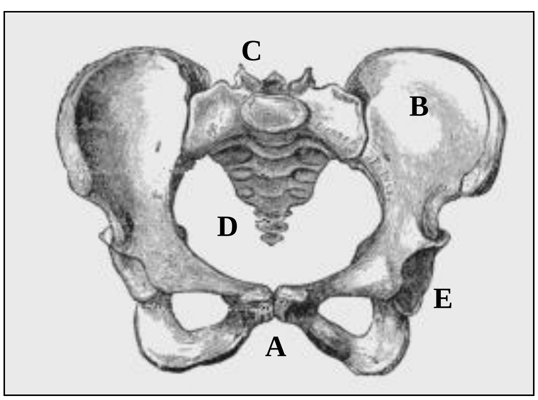
List of Tables

<u>Appendix V</u>

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The assumption that the biological anthropologist can assess age-at-death with reasonable accuracy is critical to both demographic analyses and medico-legal applications of bioarchaeology, such as forensic anthropology. In paleodemography anthropologists use age estimation to understand fertility, life expectancy, and to form hypotheses on gender roles, medical knowledge, and violence (Konigsberg and Frankenberg 1994). Forensic anthropologists use age estimation to contribute to the identification of unknown individuals and occasionally victims of violent crimes, and those techniques are exported abroad to serve in the legal persecution of genocide (Ubelaker 2008). The assumption that the methods currently used are accurate enough for these applications has been tested in many samples (e.g. Brooks 1955; Suchey 1979; Meindl et al 1985; Saunders et al 1992; Baccino et al 1991), but increasing global application of methods derived from isolated American samples continues to stimulate research advances in this field (Ubelaker 2008). Despite the general acceptance of pubic symphyseal techniques for estimating age-at-death, the scope of the factors that affect their accuracy is at issue. In particular, the complex interactions of age, sex, and ancestry and their relationship to the variation in observed pubic age have yet to be disentangled. It remains to be seen whether these factors play any greater role in age expression variation than idiosyncrasy and individualization (i.e. genetics and physical activity), which will be represented in this study by the increasing error generally found with advancing age and the decreasing accuracy of aging methods as age increases.



A: Pubic Symphysis
B: *Os Coxa* (left)
C: Sacrum
D: Coccyx
E: Acetabulum

Figure 1: The Human Pelvis

The pubic symphysis is defined as that area of jointure of the *os coxae* where they meet at the ventral, or anterior, region of the pelvis (Figure 1). With respect to forensic or bioarchaeological aging methods, the pubic symphysis refers to specifically the surface of the pubis where it is joined by cartilage to the opposite pubic surface. While this region is highly immobile, it is important to consider that this area is still a non-synovial amphiarthroidal joint that is affected by locomotion. It is the fact that the pubic symphysis displays the characteristics of development exhibited by joints that makes the pubic symphysis a likely candidate for age estimation. Its relative lack of mobility and its slightly protracted period of development make this region, in fact, ideal. This is because the pubic symphysis does display the developmental and degenerative behaviors of joints, but also because these changes can be considered age-related in a relatively predictable fashion (Todd 1920) due to the limited mobility of the joint. While other joint surfaces are affected by activity patterns during life such as carrying, walking, flexion, and strain, the relatively immobile pubic symphysis is less affected by these behaviors.

Age estimation based on more mobile joint surface wear has some evidence in favor of it (see Rissech et al 2006 for a description of acetabular aging methods); however, such methods are founded largely on degenerative processes that make age-related changes difficult to separate from the effect of activity patterns during life. Epiphyseal fusion methods, on the other hand, are based exclusively on developmental timelines rather than degenerative changes. During intrauterine, childhood, and adolescent growth periods, segments of developing bone unite at epiphyses to form bones in their adult forms. Typically, age estimation based on epiphyseal fusion is highly reliable and accurate up until the fusion of the last epiphysis, located at the medial, or sternal, clavicle (Brooks 1955). This fusion is usually completed between the ages of 22 and 30 (Scheuer and Black 2000). Epiphyseal fusion techniques lose their efficacy during the 20s and are generally useless past the age of 30.

Interestingly, the pubic symphysis spans the age gap between the epiphyseal fusion timeline and joint degeneration. Todd (1920) was the first to associate the construction and degeneration of pubic symphyseal traits such as ridges and furrows, dorsal and ventral ramparts, superior and inferior extremities, and the symphyseal rim with specific age categories, and Meindl et al (1985) observe that the immediate postepiphyseal phase of the pubic symphysis is not reached until Todd's phase seven (ages 35-39), while the degenerative phases eight through ten include individuals aged 39 through 50+. Todd's hypothesis that this skeletal region is useful for age estimation throughout life is a discovery that has led to an expansion of data and available methods that now include reliance on anatomical phases, components, statistical phases, quantitative theory models, and multiple regression analysis.

With the expansion of the body of literature on age estimation and the pubic symphysis has come a reliance on its accuracy, which needs to be continually tested using more highly varied and more representative samples. Independent variables such as race and sex also need to be tested for significance as sex- and race-specific methods continue to be developed. This project tested six methods that represented a several types of methods including the Todd (1920) anatomical phase method, the McKern-Stewart (1957) male-specific and Gilbert-McKern (1973) female-specific component methods, the Hanihara-Suzuki (1978) multiple regression and quantitative theory model method, the Suchey-Brooks (1990) male-specific and female-specific methods, and the Berg (2008) female-specific addition to the Suchey-Brooks method. All methods were tested for relative accuracy against the same randomly selected sample from the William M. Bass (WMB) Donated Collection. In addition, sex was evaluated as a significant independent variable for each method, though the number of pregnancies for each female, or parturition data, was not available and is occasionally considered relevant to aging accuracy. This donated collection was ideally suited to this project because of its size and variety, though unfortunately, race was not considered for analytical purposes because of the limited number of non-Caucasian individuals.

CHAPTER 2
LITERATURE REVIEW

The Methods

In 1920, Todd published the first work on gross anatomical changes of the pubic symphyseal region of the *os coxa* intended to provide a method of age-at-death estimation for the human skeleton. His work was based on previous research suggesting that the changes of the pubic symphysis occur throughout the life of an individual (see Todd (1920) for a historical summary of Hunter (1761), Aeby (1858), Henle (1872), and Cleland (1889)). Although most of the authors upon whose research he drew concentrated on the morphological changes of pubic symphysis, they did not relate specific bone changes to age. In order to do so, Todd ordered the skeletal series by known age and concentrated on the processes that appeared to be standard as the individual aged. These processes were then organized as phases associated with age ranges. Skeletons that did not fit the developmental standards of the time were considered pathological or deviant and thus excluded (see Gillett 1991). His methodology was based firmly in the traditional anatomical practice of simplification in order to emphasize general biological principles (Stewart 1957a). For this reason, the figures provided to illustrate each stage are modal standards representing specific cases that were felt to ideally represent each age category. These illustrations, published based on Todd's descriptions by Neumann, are provided in Appendix I.

Todd (1920) based his research on the skeletal collection of 450 individuals curated at the Western Reserve University but obtained largely from St. Louis, Missouri. This collection, now called the Hamann-Todd collection, was comprised generally of transients whose ages were estimated as the bodies were prepared for dissection, and only some were confirmed by government records (Todd 1920; Lovejoy et al 1985). He distinguished between four series within the collection for the purpose of this project: "male and female Whites, both American and foreign born, and male and female Negroes," which he classified as Negro-hybrids because of an unspecific awareness of genetic and cultural admixture (Todd 1920:287). The only series that proved sufficient in number (306) and distribution to develop a complete standard for age-related changes were white males. Todd himself acknowledged that his numbers were too small for statistical work, so he focused instead on his methodology and reasoning.

In 1955, Brooks published a critical examination of Todd's method based on demographic problems that arose when the method was applied to a Californian archaeological sample. Brooks suggested that the extreme deviants rejected by Todd were in fact representative of range of variation rather than pathology. She also found that for all ages over 20, Todd's phases consistently yielded a higher than actual age. Hanihara (1952), Krogman (1962), Meindl et al (1983), and Katz and Suchey (1986) all suggested similar findings. Further, though most of the individuals in the original sample were recorded as over the age of 40, Meindl et al (1985) reported that Todd's terminal phases actually underaged the oldest specimens. It has been suggested that both the nature of the sample itself and Todd's methodology contributed to the apparent problems with the method. Todd's focus on standards rather than variation, compounded by the poor sample documentation as evidenced by the overrepresentation of 'rounded' ages (e.g. 35, 40, 45; see Figure 2), may be responsible for its contemporary lack of utility, though both Todd (1920) and Brooks (1955) demonstrate a very high correlation between estimated and known age when this sample is applied.

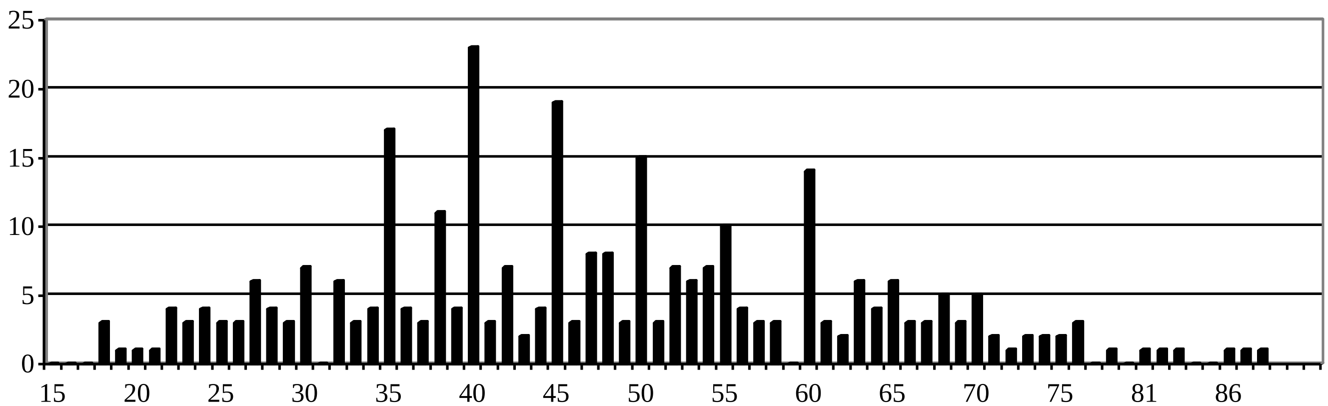


Figure 2: Sample distribution by known age (Todd 1920)

In response to both the need for an improved age estimation method and Brooks' critique, McKern and Stewart (1957) developed a three component system intended to improve on Todd's system that utilized sequential types. Their report "represents the results of extensive identification research in a thoroughly documented sample of a military population and is concerned specifically with the estimation of chronological age from the maturational status of unknown remains" (McKern and Stewart 1957:11). It is based on a sample of 450 skeletonized and identified U.S. war dead that were being repatriated from North Korea in 1954 (McKern and Stewart 1957). Their documentation was more thorough than Todd's, and included age, state or territory of origin, race, and ethnic background, all verified by military records.

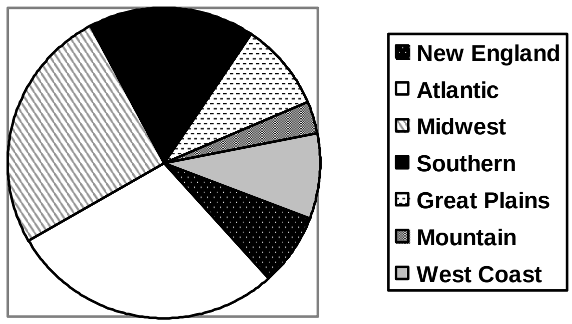


Figure 3: Distribution by region of origin (Stewart and McKern 1957)

Notably, they were also concerned with the specific effects of environmental conditions such as malnutrition and stress, though they ultimately determined that those individuals that had been held as prisoners of war (133 individuals, the majority held from 3-6 months) did not demonstrate significantly more or less skeletal indicators of stress measured by cranial and postcranial osteoporoses than their peers killed in action (Stewart 1957). Geographically, the distribution is similar to that of the American population in general (Figure 3); however, age, sex, and race are heavily biased toward young male Caucasians consistent with the demographics of the military at the time (Figure 4).

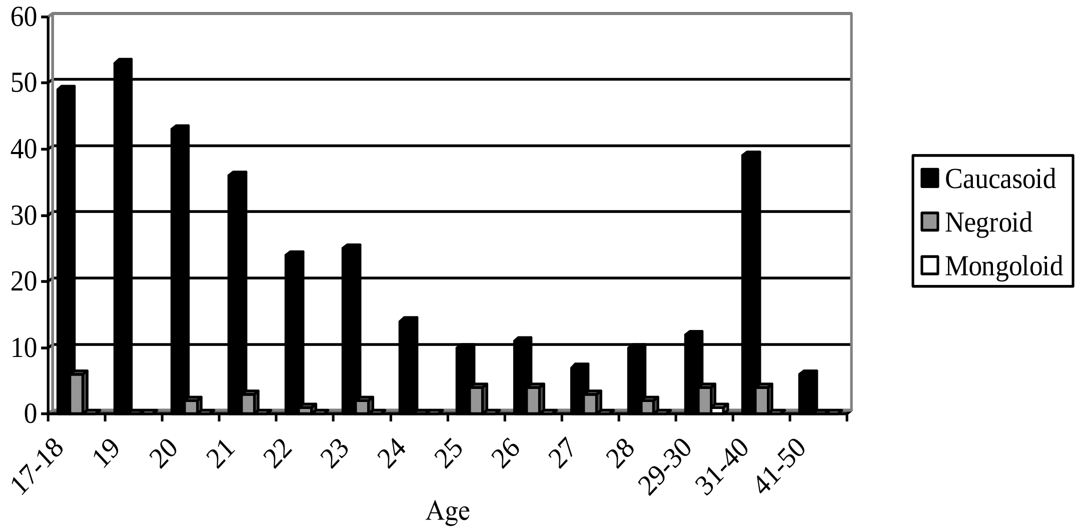


Figure 4: Sample distribution by age and race (McKern and Stewart 1957)

McKern and Stewart used a derivative of a somatotype formula (Sheldon 1940) which uses three components of seven grades each, in place of Todd's system of successive phases. They suggested that the use of a formula would be ideal for comparative purposes, but would also allow the observer to analyze the composition of a structure and the reader to visualize what the original structure looked like (Stewart 1957). It also provided for the possibility that the components might vary independently. In order to establish components that could be graded, they analyzed the age-sensitive diagnostic features originally described by Todd and grouped them as the ventral rampart, the dorsal plateau, and the symphyseal rim, each of which could then be scored individually on a scale of 0-6, recombined, and then associated with an age range. The illustrations McKern and Stewart published of the stages of each component intended to be used in conjunction with this method are provided in Appendix I. By using a component method, McKern and Stewart hoped to encompass the individual variability in pubic symphyseal morphology that was lost in Todd's standards (McKern and Stewart 1957).

In subsequent tests, the McKern-Stewart method did not prove useful for traditional, non-military samples because its upper bound is very low (36+), and the original database is heavily concentrated on the very young. Klepinger et al (1992) noted, however, that though the system may be flawed because of the restricted database on which it was built, the McKern-Stewart standard deviations were unrealistically constricted, which inflated the comparatively poor performance of this method.

Todd (1921:37) had reported that “[sex-related] features have but the most meager influence upon the estimate of sex, and taken in a single case, would be of very doubtful value.” While his female sample size may have been far to small (69 individuals) to draw such a conclusion, Todd (1921), Brooks (1955), and McKern and Stewart (1957) all indicated that the methods in existence were less reliable for females than for males, which may be at least partially attributed to the fact that all methods at the time had been generated based on male samples. In response to this bias, Gilbert and McKern (1973) published a three-component system similar to the McKern-Stewart method that was female-specific and based on a sample of 103 American females aged 13-57, concluding that “females are absolutely different from males in the rate and locality of age-related metamorphic changes in the *os pubis*” (1973:31). In contrast to Todd’s determination, Gilbert and McKern (1973) found considerable differences in the timing of maturational processes; for example, Todd (1921) stated that the flattening of the dorsal demiface occurred approximately two to three years earlier in females than in males, whereas Gilbert and McKern (1973) found that the dorsal flattening occurred by ages 25-28 in females, but not usually in males until 35+ years. The proposed alternate morphology for females was illustrated to provide assistance in age estimation; these illustrations are provided in Appendix I.

Gilbert (1973:40) later reported that further testing indicated that “the female standard yielded estimates within useful limits in all age ranges.” Gilbert and McKern (1973) reported no regular metamorphic activity in individuals beyond 55 years; thus, like Todd’s and McKern-Stewart’s methods, the Gilbert-McKern method is not useful for aging the elderly. Klepinger et al (1992) confirmed that the average absolute deviation of the true age from the mean age was decreased by this method in comparison to the Suchey-Brooks system for females; however, Suchey (1979) repeated the test using 11 known pelves and the individual assessments of 23 trained observers and found that only 51% of the assessments yielded age ranges which included the known age of each specimen. She questioned whether this error was due to the method itself, which may be unclear in its phase delineations, or to the nature of the variability present in female pubic symphyseal morphology. When Gillett (1991) compared demographic profiles of the same Central California shell mound site using Todd, Gilbert-McKern, and Suchey-Brooks, he found that all methods appeared to skew the female data towards the upper decade. This apparently inherent variability in the female *os coxa* is frequently attributed to the presumed skeletal trauma of childbirth (see Gilbert and McKern (1973)); though Todd (1920) and Hoppa (2000) refuted this (the latter also suggested hormonal activity and difference in gait mechanics as possible contributors). Regardless of the source, testing throughout the decades has repeatedly indicated that the appearance of the female pubic symphysis is more likely to seem older than it actually is.

In 1978, Hanihara and Suzuki developed a seven-term multiple regression model of pubic symphyseal aging based on a Japanese sample of 70 pairs of pubic bones from individuals aged 18-38. This system was inspired by earlier methods: the scoring elements were adapted from Todd's phases, and the treatment of the scores diverges slightly from the method developed by McKern and Stewart. Once scores for each of the seven morphological features of the pubic symphysis are obtained, they are used as raw data for a multivariate statistical analysis, whereby the skeletal age can be estimated by calculating a simple linear function from the multiple regression analysis (MRA). In addition, Hanihara and Suzuki (1978) performed a quantification theory model I analysis (QMI) which provided a slightly higher correlation coefficient with known age. Since Todd (1921) reported no clear differences between the sexes in age changes of the pubic symphysis, Hanihara and Suzuki (1978) combined the sexes in this study, though they noted that future work should strive to separate the samples by sex to improve accuracy.

Hanihara and Suzuki (1978) posited that age differences were simply not considerable or reliable enough in the older middle ages to warrant inclusion in their study. For this reason, they limited the age distribution of the sample. In a test of this system, Meindl et al (1985) observed that though the system was based on a limited sample in terms of total number, and sex and age distribution, the accuracy was generally better than that of other component systems in the 20 to 40-year age range. Hanihara and Suzuki (1978) found that the QMI was slightly more reliable than the MRA for samples in age groups one and two (18-30 years), but that the two are equally reliable for age group three (30-38 years).

In 1986, Katz and Suchey published a new method based on the largest and most diverse sample to date that had come from the Department of Coroner, County of Los Angeles from autopsies performed during the summer of 1977. The series of 1225 individuals consisted of 739 males and 273 females between the ages of 14 and 99, all of whom were accompanied by legal documentation of birth and death dates. Parity information was also obtained for the 273 females. They made serious attempts to ensure the ancestral, sexual, occupational, and socioeconomic heterogeneity of their sample to ensure universal applicability. Further, no individuals were deleted from the sample because their morphological traits were not consistent with the stated age. The resulting Suchey-Brooks method was ultimately based on a sample composed of individuals born throughout the United States and 31 foreign countries, including Europe, South America, and Asia. The geographic origin of each individual was determined by appearance only in the autopsy room, using skin color, hair color, hair form, nose form, amount of alveolar prognathism, lip form, amount of body hair, and the prominence of zygoma. To describe a Mexican category, Katz and Suchey also examined incisors for shovel-shaping and utilized records of birthplace.

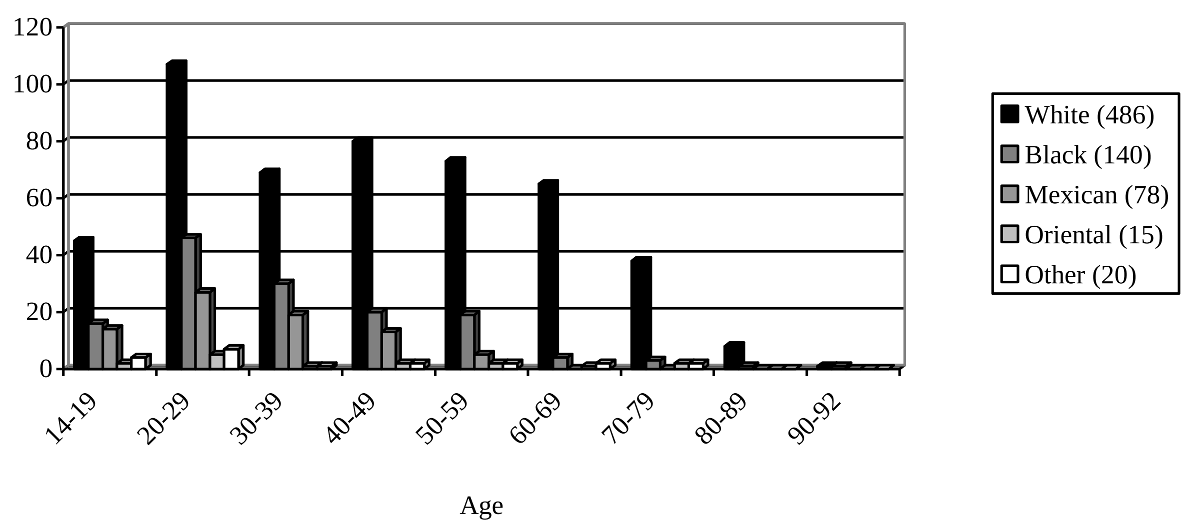


Figure 5: Male sample distribution by age and race (Katz and Suchey 1986)

In this method, Todd’s ten phases were consolidated into a six-phase system based on the problems and data reported by Brooks (1955). Categories were collapsed because Katz and Suchey (1986) found that observers could not consistently distinguish between them. The authors found that all methods performed poorly when the entire sample is used, and improved results were achieved consistently by eliminating older individuals, though they recommended the modified six-phase system for practical implementation. Katz and Suchey (1986) argued that the component systems utilized by McKern and Stewart and Gilbert and McKern were overly complex because the components do not in fact vary independently. Katz employed linear regression methods to establish appropriate chronological phase ranges, and found that small improvements in estimated error could be obtained by using more complex multivariate procedures. Ultimately, the simpler six-phase procedure proved to be the most applicable and successful. This system was originally published in conjunction with illustrations for each phase, but later incorporated a cast system intended to improve accuracy of phase assignments in the field.

Baccino et al (1993) found that the Suchey-Brooks six-phase system offers advantages over the alternatives, especially with regard to the availability of cast examples, though Kimmerle et al (2008b) found more observer variation with this method than with the Todd method, suggesting that its application is possibly unclear. Galera et al (1995) had also found substantial inter-observer variation in the Suchey-Brooks application, but noted that this method remained more reliable than the Todd method. With regard to the sample itself, Klepinger et al (1992) suggested that tests of American black and white males may be inherently biased because of the generally younger ages of the black individuals in the autopsy samples, which created some questions concerning the later Katz and Suchey racial refinements. When the mean age for each group is calculated using the average of the age ranges, the Caucasian sample is proven to be significantly older on average than either the African-American or Mexican samples (Table 1). Given the size of the samples for each group (represented by *n* in Table 1), it is unclear how influential this discrepancy is either on the age studies themselves or on the methods published by Katz and Suchey (1986).

Table 1: Mean age by group affiliation (Katz and Suchey 1986)

Group	***n***	**Age range**	**Mean**
Caucasian	486	14-92	43.36
Black	140	14-92	35.78
Mexican	78	14-59	30.76

Berg (2008) further revised the Todd method by adding a seventh phase to the Suchey-Brooks modification for females only. He reported that this inclusion resulted in *r*-values consistently better than those reported for the Suchey-Brooks method, even though the sample included not only young individuals, but also large quantities of older females. Though most researchers state that estimating age-at-death for the elderly is far less reliable than for younger individuals (see Hanihara and Suzuki 1978, Katz and Suchey 1986, Meindl 1985, and Klepinger et al 1992), Berg found an overall high degree of accuracy using the new seriation. Kemkes-Grottenthaler (1996) had previously attributed the increasing variation with age to a variety of non age-related determinants, including population-specific factors, pathological conditions, and environmental parameters that influence bone structure; however, Rissech et al (2004) reported that only the more severe manifestations of specific joint syndromes including ankylosing spondylitis, psoriatic arthritis, and Reiter's syndrome might obstruct the reliability of age estimation at this articulation.

In terms of analysis, the focus of this project is on the tested bias and inaccuracy of each method applied to the WMB collection. While inaccuracy refers to the actual deviance of the estimated age from the actual age, bias refers to directional inaccuracy, or the method's tendency to overage or underage the tested individual. Because all methods employed in this study, with the exception of the Hanihara-Suzuki regression methods, are range methods, accuracy was tested by artificially creating single point estimates for each range to compare with actual recorded age. Since a major focus in the development of age estimation methods has been "coverage" (Konigsburg et al 2008), which refers to the adherence of a method's practical results to published expectations, the study includes coverage data but focuses on mean bias and inaccuracy. For example, if a method has 95% coverage, then 95% of the test individuals displaying the characteristics of a given phase must have ages that lie within the stated age range of that phase, but the tested coverage could not be used to derive bias or absolute inaccuracy. The difficulty with the concept of coverage in paleodemographic but especially forensic application is how much and how big: clearly 100% coverage would be ideal, as this would imply that all unknown individuals would be placed in an appropriate range, but if this range includes ages 30-100, then it is practically unwieldy and statistically useless. Analytically speaking, this study is geared more towards investigating the statistical contribution of certain isolated variables to relative tested accuracy.

Regression and Bayesian analyses have been used as the current statistical methods of choice for testing age estimation in many capacities (Lucy et al 1996, Aykroyd et al 1999), but there are some limitations to both that impinged on their success in this particular application. Lucy et al (1996), for example, discussed a series of assumptions about the nature of data that must be met for regression analysis to be useful. A regression analysis requires that variables vary continuously with age, meaning that a variable must theoretically be capable of adopting an infinite number of values. In contrast, the phase and component methods that are discussed in this paper use ordinal rather than continuous variables. The fact that the single point data for each phase is artificially contrived for the sake of comparison tampers with the integrity of the regression correlation intended to model the relationship between two continuous data sets, actual age and estimated age. Instead, bias by age was depicted graphically for each method to provide a visual approximation of the correlation of the point estimates and actual age. Lucy et al (1996) observe that though there is no real reason a regression analysis shouldn't be applied for categorical or ordinal data, assumptions concerning data should be checked, and other techniques should be considered. Similarly, Konigsburg and Frankenberg (1994) stated that regressing actual age on an indicator requires the assumption that the sample under study displays a similar age-at-death distribution as the reference sample. With respect to this study the distribution is in all cases significantly different for a variety of contextual reasons related to the individual collections that form the basis of each method.

Konigsburg et al (2008) argued that age estimation for the purpose of evidence should always have a Bayesian underpinning, but the application of Bayes' theorem hinges on having a prior distribution for the estimate. This means that one must have some knowledge of prior age-at-death distributions relevant to the tested individual in order to make a reasonable guess at possible age. Kimmerle et al (2008a) used a Bayesian analysis to explore the likelihood of age parameters biased in the direction of the reference sample which may contribute to poor reliability when methods are applied to collections presumed different from the original sample such as in Kosovo and Croatia.

They concluded that the most accurate parameters are achieved when revising the calibration for estimating age among males and females. While this type of analysis is achievable using data obtained from records held on the collection under study, this statistical method would be overly complex and immaterial for a test focused on the application of aging methods to an unknown sample.

A final type of statistical analysis used in methods of age estimation has been transition analysis proposed by Boldsen et al (1992) and implemented, for example, by Konigsberg et al (2008) and Berg (2008). The purpose of utilizing this method is to improve on linear regression, which overages younger individuals and overages the elderly, by focusing on the transitions between phases and providing a point estimate for ages at transition. As with Bayesian analysis this method can be applied to unknown individuals but only in cases where age distribution information is obtained about the sample to which the individual pertains.

The statistics used in this study focused more strictly on the difference between application for male and female samples and the practicality and relative accuracy of each method. Klepinger et al (1992) published a similar study that measured performance by comparing absolute deviation of true age from interval means, and measured the frequencies of true age falling within one and two standard deviations above and below stated means. These tests are reproduced when analyzing the Berg 7^{th} phase addition to the Suchey-Brooks method, and to a limited extent for the other methods in this study. Meindl et al (1983) also tested bias and inaccuracy of several age estimation systems including the pubic symphysis. They reported a marked tendency of all tested methods to underage in the fifth and sixth decades of life, which can be more closely examined using the WMB collection due to its older average age.

Ancestry

Concern over the influence of ancestry on skeletal aging has been a consistent theme in age studies, and the specific nature of this concern has changed as the ideas of ancestry and race have been conceptually revised. In 1920, for example, Todd examined individuals from different “racial stock” in order to define differences between them in his aging studies. He observed only minor differences that he concluded were largely unimportant, though subsequent studies have suggested that there are in fact differences that could be interpreted along racial lines: Sinha and Gupta (1995) observed significant differences in the mean age of phases in their analysis of McKern and Stewart's method in a sample of males from India. Komar (2003) also criticized the universal application of methods derived from American samples, reporting that only 20% of the individuals greater than 50 years old were aged accurately (even using large age range estimates) in a Bosnian forensic population.

Hanihara (1952:255) found that Todd's phases tended to overage many individuals, but concluded that "the age changes of the Japanese people are, generally speaking, two to three years earlier," suggesting that American methods may in fact need to be revised for other populations. In contrast, Sakaue (2006) reported that though the Suchey-Brooks system was based on a collection of pubic bones from autopsied individuals from the United States, this system was applicable to contemporary Japanese skeletal material. The results of this test were similar to those reported by Hanihara (1952) with mean ages of the Japanese and Suchey-Brooks series differing by less than three years, but Sakaue (2006) concluded that application of the method was therefore comparatively reliable. Kimmerle et al (2008a) found significant variation between American and Balkan female populations, but no difference between corresponding male populations when the Suchey-Brooks method was applied. In 1985, Meindl et al performed a three-way factorial analysis of variance (expected (bias) = mean + race + sex + age + interactions) and found no race-sex combinations that produced significant bias in age determination. They also found that the differences in error in age estimation between races were non-significant, based on a sample from the Hamann-Todd collection.

Conceptual changes in the understanding of race in the United States have led physical anthropology in general to privilege a more population-based theory of ancestry as opposed to the typological and socially problematic idea of race. Ancestry is based on the foundation of environmental adaptation that has produced the physical differences observed between global human populations, but is unrelated to their social differences. Kemkes-Grottenthaler (1996) cautioned that perceived population differences (i.e. racial differences) in pubic symphyseal aging methods may be the result of extrinsic factors such as material culture (e.g. health care access, nutrition, or physical activity types and levels) which affect bone density or degeneration. Katz and Suchey (1989) did find significant differences in age across racial groups – African-Americans and Mexicans with advanced pubic symphyseal patterns tended to have lower ages than Caucasians – but the authors acknowledged that causality could not be attributed to the variable of race itself. Instead they suggested that "genetic factors and/or environmental variables such as diet, alcoholism, or drug abuse" might be involved in the apparent differences in appearance (Katz and Suchey (1989:167). It is important to reference again the highly significant differences between average ages for each category in this autopsy sample which may have affected the integrity of the comparison. In addition, poor accuracy results when aging methods are globally applied might be related to previously recognized problems with the specific methods: it is generally agreed that aging females is less reliable than aging males; that individuals over the age of 50 are difficult to age with an appreciable degree of accuracy; and that the McKern-Stewart reference population is very limited and unable to provide a reliable basis for aging the elderly or even the middle-aged.

Sex

With respect to the female pelvis, the trauma of pregnancy has repeatedly surfaced as a potential contributor to its apparently higher variability (Todd 1921; Stewart 1957; Gilbert and McKern 1973; Suchey 1979; Suchey et al 1988). Todd (1921; 1923) reported no evidence that this might be the case in his own sample and in his review of earlier sources, though he found significant physiological joint changes in pregnant experimental rodents, particularly guinea pigs, during related research. Hunter (1761, in Todd 1921) stated that there was no real difference in the region of the symphysis pubis between the pelvis of a parturient and that of a non-pregnant woman. In addition, Aeby (1858, in Todd 1921:40) argued that "no increase occurs in the actual distance between the pubic bones during pregnancy but that there is a softening and consequent extensibility in the ligaments themselves," which Todd argued should not leave any permanent stamp upon the skeleton. Hoppa (2000) later incorporated parturition data in tests of aging methods but concluded that obtaining these data did not improve overall accuracy.

In contrast, Stewart (1957) suggested that the dorsal aspect of the female pubis was in fact altered by pregnancy, and both he and Angel (1969) described the appearance of these changes. They argued that fetal development may cause the inter-pubic ligaments to be pulled to the point of hemorrhage, which led to the appearance of pits or grooves on the dorsal surface of the pubis. According to Putschar (1931; 1976) and Heyman and Lundquist (1932), the pubes separate at the symphysis as much as 0.5-10 mm during the last weeks of pregnancy, and the inter-pubic ligaments attach more and more laterally from the symphysis with each successive pregnancy. This would support the claim that pregnancy, especially multiple pregnancies, would leave permanent indications on the region of the pubic symphysis; however, Gilbert and McKern (1973) examined 140 cases of known parity and found that the number of pregnancies could not be reliably determined by the appearance of the *os pubis*. These cases included involvement of the dorsal aspect of the symphyseal rim and the dorsal demiface, and the authors observed that apparent parity trauma could cause an individual to appear older. They argue that a sample of nulliparous individuals would yield much smaller standard deviations than those they achieved in their study (Gilbert and McKern 1973).

Hermann and Bergfelder (1977) repeated a similar study of 49 individuals of known parity and found no criteria useful for clear forensic diagnoses of childbirth, but they did observe a correlation between parturition and the morphology of grooving on the posterior pubic cortex. Walde (1962) confirmed that among the typical obstetrical complains is pelvic pain that he associated with the softening processes of the pubic symphysis, and referenced Loeschke’s (1912) findings that connective tissue hypertrophy, cartilaginous changes, and vascularization with hermorrhage are related to traumatic symphyseal fissures in pregnant women. Although Walde did not relate this to specific osteological changes, it seems probable that this joint implication in pregnancy-related hormonal changes could in some way impact the bone itself. Regardless of the differences in opinion, it seems apparent that individuals progress through variously defined age phases at rates subject to both environmental and genetic constraints that remain poorly understood (Berg 2008). Ultimately, parity seems to be a statistically irrelevant if not practically unimportant when it comes to age estimation from the pubic symphysis, a fact simplified by the unavailability of this data for this sample.

Meindl et al (1985) also suggested that for females, absolute pubis size plays a significant role in the accuracy of age estimation. Washburn (1942) did not relate the dimensions of male and female pubes to age estimation, but he did observe that while the length of the ischium is roughly proportional, females have proportionately longer pubic bones than males; thus, it is possible that the structural differences are related to different morphological appearances. At this time, sufficient data has yet to be compiled that would be useful for correcting age estimation for either pubis size or parity. The use of sex as a potential influencing factor on the timeline or route of skeletal change with age in this study will serve to provide direction for further research into this topic, if not provide the background in skeletal aging processes that is necessary to improve upon current methods.

CHAPTER 3
MATERIALS AND METHODS

The Sample

Three hundred and ninety-six individuals of known age-at-death, sex, and ancestry were randomly selected from The University of Tennessee-Knoxville William M. Bass Donated Skeletal Collection housed at the Forensic Anthropology Center (see Figures 6 and 7 for sample distribution by ancestry, sex, and age).

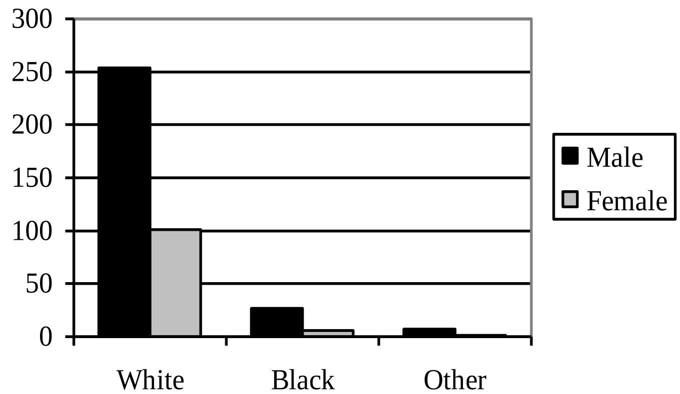


Figure 6: WMB sample distribution by sex and ancestry (n=396)

This collection was particularly well-suited to this project because of its well-documented nature and the fact that no age estimation methods used in this study were developed based on this sample with the exception of the Berg (2008) revision. A blind study was used to evaluate age-at-death using each of five methods, including the original Todd (1920) method, the McKern-Stewart method (1957), the Hanihara-Suzuki regression method (1978), the Gilbert-McKern method (1983), and the Suchey-Brooks male and female revised methods (Brooks and Suchey 1990). Where the Suchey-Brooks method was used, the seventh-phase addition proposed by Berg (2008) was also evaluated concurrently. During analysis, the Suchey-Brooks method's accuracy was interpreted both with the Berg seventh-phase addition and without it.

Since the Todd (1920) method was concerned largely with identifying and describing ideal anatomical types, his descriptions will be given preference, although illustrations of modal types dictated by Todd (1920) were provided by Neumann (*in* McKern and Stewart (1957)) and were also available during evaluation. These illustrations are available in Appendix I for reference. The McKern-Stewart (1957), Hanihara-Suzuki (1978) regression, and Gilbert-McKern (1983) methods were used with the accompanying illustrations and photographs for reference. The comparative casts for the Suchey-Brooks (1990) phases will be used as this technique represents an integral component of the application of this method.

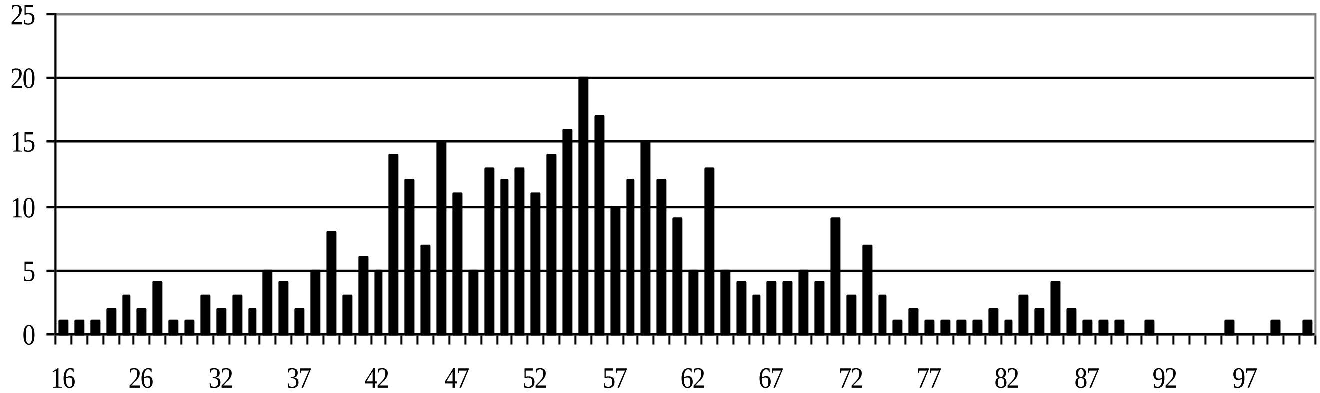


Figure 7: WMB sample distribution by known age (n=396)

Berg's (2008) illustrations will also be incorporated in this review of the sample. All illustrations, plates, and photographs used during this evaluation are found in Appendix I.

Before beginning this test a colleague's help was enlisted to prepare the sample and obscure the identifying information on each individual's container to prevent it from influencing the data collection. He was instructed to select individuals without regard to age, sex, or ancestry, but to exclude individuals with bilateral sacroiliac fusion or postmortem damage to the pubic symphysis, which would obstruct evaluation. Cases with unilateral or bilateral joint replacement of the hip were kept as they represent a type of contemporary normal variation in forensic anthropology. An independent comparison of the six techniques was best served if the age of each individual could be estimated using all six methods; however, it was important to prevent the results of one method from informing another. For this reason, the sample was reviewed in its entirety six separate times, and it was expected that intra-observer variation would be controlled by completing all age estimations for one method before introducing subsequent methods. It was also reasonable to expect that the observation of 396 individual pubes would provide sufficient time to prevent the memory of one set of results from influencing subsequent reviews of the sample. Data record sheets included descriptions of phases or components used in all cases in conjunction with illustrations or casts where appropriate, but ages were excluded to focus the examination on the pubic symphyseal structures (Appendix II). The data was then entered into five separate electronic spreadsheets, where it was organized by case number for each individual.

The five data sets were united and associated with the recorded age and sex of each individual. A sixth set was created to separate the Suchey-Brooks (1990) results from the Berg (2008) 7^{th} phase addition. All data sets were entered in SPSS 16.0 statistics software package for Windows, Student Edition. Standard deviations and mean actual ages for each phase were calculated for all methods for comparison to the original data and subsequent test results from other collections (Appendix IV). In order to facilitate comparison and analysis, the techniques that produce estimates as age ranges were converted into single ages derived from the mid-point of the range. Both the signed difference between estimated (phase midpoint) and actual age (bias) and the absolute difference (inaccuracy) were calculated for each of the methods against the recorded age-at-death of each individual. Actual age and sex were termed independent variables, and bias, inaccuracy, and phase determination were termed dependent variables (Table 2). In addition to providing mean absolute deviance from phase midpoints and mean bias for both sexes and a combined sample where appropriate, a visual graphic was given to illustrate the trend in bias across age in order to expand upon the conclusions drawn by the means.

Table 2: Dependent and independent test variables

Independent	Independent	Dependent	Dependent	Dependent
Actual age	Sex	Bias	Inaccuracy	Phase

A Levene's test for equality of variance was employed using SPSS in order to test for homogeneity of variance for bias and inaccuracy across sex categories. Race categories were excluded due to the limited sample diversity in terms of reported ancestry. Once equality of variance was established, an independent samples test for equality of means was used to determine whether mean bias or inaccuracy across sex categories were significantly different. Mean bias and inaccuracy were used in favor of regression or Bayesian analysis for specific reasons related to reference and test sample distribution differences and the exclusive focus of this study on practicality and efficacy of method application to an unknown individual.

Limitations

A major concern regarding skeletal age estimation is that many agree that multifactorial aging methods that incorporate as many indicators of age as are available offer better results than those utilizing single indicators (Lovejoy et al 1985; Bedford et al 1993). Rogers and Saunders (1994) recommend a maximum of six methods used concurrently. Certainly more extensive analyses are preferable and improve overall statistical accuracy, but relevance for application is based on practicality in addition to accuracy, and multifactorial methods often call for complicated analyses that are impractical in the field or in large-scale sampling. Macroscopic methods are preferred over microscopic reasons for similar reasons, even though they may not be as accurate (Aiello and Molleson 1993).

Further potential limitations to this study exist within the methods themselves. Owings Webb and Suchey (1985) suggested that age standards developed using historical osteological collections "do not have any relevance for living populations because of today's widespread use of oral contraceptives and vitamin D supplements," and because of "secular acceleration," which has "manifested itself in a general acceleration of growth and development and has changed the onset of such life markers as menarche and menopause" (Kemkes-Grottenthaler 1996:281). As all reference samples are by their very nature historical due to the inconvenience of time, and the fact that methods developed even in the 1920s remain in use with acceptable tests of accuracy, this hypothesis cannot be considered a limiting factor, though it will be interesting in further study to compare the accuracy of age estimation on the WMB collection to results from tests of the older Terry and Hamann-Todd collections.

CHAPTER 4
RESULTS

Todd (1920)

One of the remarkable differences between the Hamann-Todd collection which was used to create the Todd method and the WMB collection used in this test was the fact that the WMB collection is entirely donated. This means that not only is the information about the collection available for research more accurate and more complete, but the socioeconomic statuses of the individuals are significantly different. The Hamann-Todd collection is comprised generally of transients and individuals left in hospital morgues, which suggests lower socioeconomic class. Higher socioeconomic status is consistently related to better health care and nutrition, which is highly relevant to bone health (Elliot et al 1996; Ashby et al 2007). Because of this difference the WMB collection may be a better analog for the general American population, but the socioeconomic facts of the two collections should be taken into account when applying the Todd method to a collection with higher average socioeconomic status.

The Todd ten-phase system was applicable for all 396 individuals in the sample. Though he did publish an adjusted system for females only in 1921, the age diversity and size of that sample has led the original system, though indicated only for Caucasian males, to be applied universally. Ultimately, 60% of the tested individuals' evaluated age ranges contained the actual age of that individual, or method coverage was 60%. The tested ranges and means for each phase are presented in Table 3 with the original ranges and means given for each phase by Todd (Table 3). Because a midpoint for phase ten (age range 50+) could not be calculated, the average bias and inaccuracy of this method were calculated based on phases one through nine (n=148). The average bias for these combined phases is -3.81 and the average inaccuracy is 8.45.

Table 3: Todd (1920) WMB tested phase ranges, means, and standard deviations

Todd (1920)	n	% in range	WMB Test Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Phase 1 (18-19)	1	0%	16	16	0	18	0.0
Phase 2 (20-21)	1	100%	20	20	0	20.5	0.7
Phase 3 (22-27)	2	50%	25-38	31.5	9.2	23.71	1.7
Phase 4 (25-26)	2	0%	24-27	25.5	2.1	25.6	0.6
Phase 5 (27-30)	3	0%	24-38	29.0	7.8	29.33	2.7
Phase 6 (30-35)	10	20%	25-71	45.2	13.7	33.56	2.2
Phase 7 (35-39)	12	50%	29-64	39.6	9.4	36.91	2.8
Phase 8 (39-44)	26	32%	26-74	45.6	11.2	41.07	5.0
Phase 9 (45-50)	90	20%	35-89	50.7	10.3	52.62	9.3
Phase 10 (50+)	249	81%	35-99	62.8	12.7	63.56	11.4

The fact that the ten-phase Todd system was not created on a statistical basis proved disadvantageous primarily because the age ranges for each anatomical phase are unrealistically constricted. In addition, the focus on the younger age ranges (18 to approximately 30) seems inappropriate not only for this sample, whose average age is 58.5 years, but for the general American population, whose average life expectancy is 78.1 years as recorded in 2006 (Centers for Disease Control and Prevention 2008). The poorer performance of this method as age progresses past 50 is illustrated in Figure 7; however, the performance of the method in the younger age ranges is difficult to evaluate with such small numbers of individuals represented by these phases in this sample. It is very clear from the data presented in Figure 7 that the results of this test are in line with Pal and Tamankar (1983), Mendl et al (1985), and Katz and Suchey’s (1986) findings that Todd’s method underages older individuals. These results could be attributed to the limitations of the original sample discussed previously or to the limitations inherent in Todd’s methodology itself.

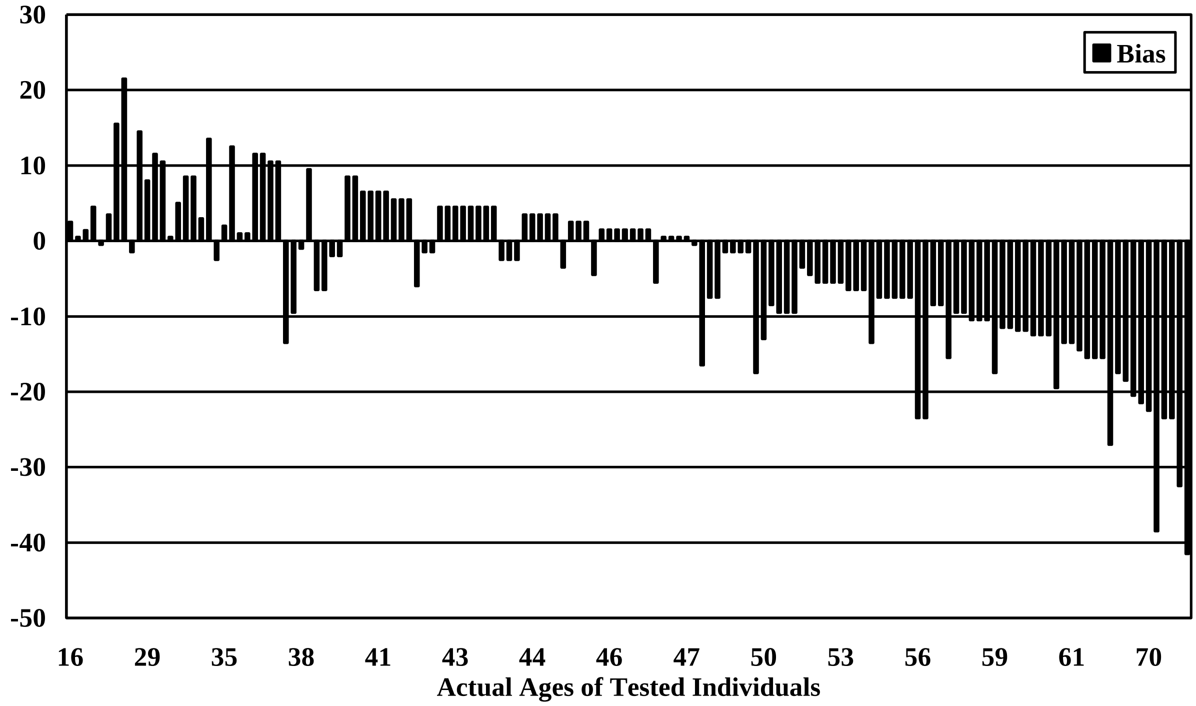


Figure 8: Tested bias and actual age for Todd (1920), WMB sample

Bias and inaccuracy could only be calculated for those scores in each method that provided a closed age range. Here, the means for both bias and inaccuracy are different for males and females given scores from 1-9: on average, ages provided by the Todd method for females are more accurate than the ages provided for males. Although this method underages both sexes (Table 4), the ages provided for males were more negatively biased. In absolute terms, inaccuracy is more extreme for males than it is for females in this sample.

Table 4: Descriptive statistics for Todd phases 1-9 for males and females

Group Statistics					
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Bias	Female	23	-2.543	7.5767	1.5799
	Male	125	-3.904	10.8922	.9742
Inaccuracy	Female	23	6.478	4.7109	.9823
	Male	125	8.816	7.7638	.6944

The Levene's test for equality of variances indicates whether males and females as two discrete groups have approximately equal variance on the dependent variable. For both bias and inaccuracy, the probability that these numbers could occur by chance is less than 5% (p > .05), which means that the two variances are not significantly different, so equal variances can be assumed for an independent samples test. This second test was used to evaluate the hypothesis that the means of inaccuracy and bias are equal for male and female groups. The variables ‘bias’ and ‘inaccuracy’ were compared between levels one and two of the grouping variable ‘sex.’ The results of this test suggest that the Todd method resulted in no significant difference in mean bias and mean inaccuracy between males and females (Table 1, Appendix V).

McKern-Stewart (1957)

The McKern-Stewart three-component system was developed using a sample of American war dead from the Korean War and is therefore recommended for males only (n=288). In this test the method’s performance was evaluated for both males and females combined in addition to males alone to establish a basis for comparison. In order to estimate an age range for an individual using this system, the sum of the independent scores for each component is associated with the published age ranges. The tested means and ranges for each score is provided in Table 6 (males and females) and Table 7 (males only) with the means and ranges given by McKern and Stewart. When both males and females were sampled together, 82.2% of the individuals were placed in appropriate age ranges, and for males alone, 77.8% of the individuals were placed in appropriate age ranges. As with the Todd system, the ultimate and penultimate age ranges, 29+ and 36+, were not conducive to calculating bias and inaccuracy based on midpoints, therefore these figures were calculated only for all scores from four through thirteen. There were no individuals that scored less than four within the WMB sample. For all scores four through thirteen (n=74), the average bias is -14.6 and the average inaccuracy is 15.4. For females alone (n=9), these figures are -16.2 and 16.3, respectively; for males alone (n=65), these figures are -13.2 and 13.2, respectively.

Table 5: McKern-Stewart (1957) tested phase ranges, means, and standard deviations for males and females

McKern/Stewart (1957) Males and females	n	% in range	WMB Test Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 0 (>17)	0	--	--	--	--	--	--
Score 1-2 (17-20)	0	--	--	--	--	--	--
Score 3 (18-21)	0	--	--	--	--	--	--
Score 4-5 (18-23)	4	25%	20-38	27.0	7.7	20.84	1.13
Score 6-7 (20-24)	4	25%	16-54	35.8	18.6	22.42	0.99
Score 8-9 (22-28)	6	33%	24-59	37.5	13.6	24.14	1.93
Score 10 (23-28)	7	14%	23-71	41.0	15.4	26.05	1.87
Score 11-13 (23-39)	53	23%	26-65	46.4	9.2	29.18	3.33
Score 14 (29+)	49	100%	35-88	53.6	11.7	35.84	3.89
Score 15 (36+)	273	95%	26-99	60.7	13.9	41.00	6.22

Table 6: McKern-Stewart (1957) tested phase ranges, means, and standard deviations for males only

McKern/Stewart (1957) Males only	n	% in range	WMB Test Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 0 (>17)	0	--	--	--	--	--	--
Score 1-2 (17-20)	0	--	--	--	--	--	--
Score 3 (18-21)	0	--	--	--	--	--	--
Score 4-5 (18-23)	3	0%	25-38	29.3	7.5	20.84	1.13
Score 6-7 (20-24)	2	0%	16-49	32.5	32.3	22.42	0.99
Score 8-9 (22-28)	3	67%	24-32	27.0	4.4	24.14	1.93
Score 10 (23-28)	6	16%	23-71	41.3	16.8	26.05	1.87
Score 11-13 (23-39)	51	23.5%	26-65	46.4	8.9	29.18	3.33
Score 14 (29+)	39	100%	43-88	53.4	11.1	35.84	3.89
Score 15 (36+)	183	92.9%	26-96	55.1	12.3	41.0	6.22

Although both male- and female-only sub-samples used to calculate bias and inaccuracy are extremely small, it is interesting to note that almost the entirety of the inaccuracy in both cases is due to underaging. This is consistent with much of the criticism leveled at this method in testing subsequent to its publication (see Sinha and Gupta (1995) and Katz and Suchey (1986)) and the most frequently attributed cause is the original highly specific sample of Korean war dead on which this method was based. It is unsurprising both that this method performed relatively poorly for advanced ages and for females, which were not encompassed by the scope of the original study.

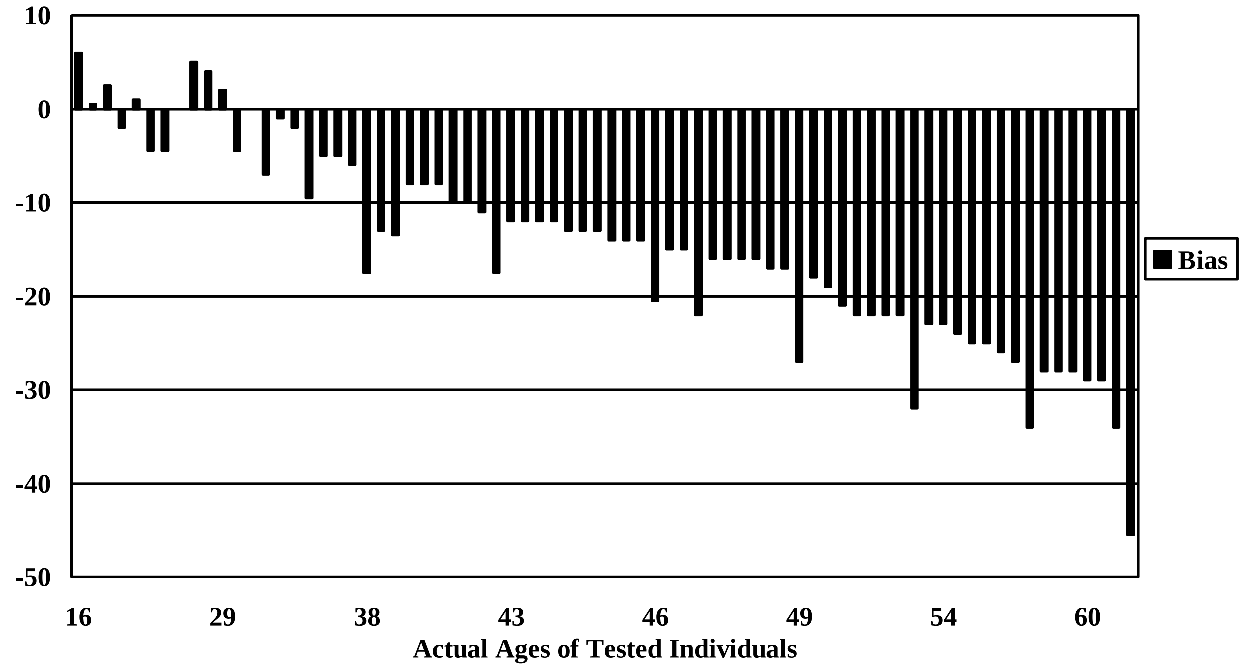


Figure 9: Tested bias and actual age for McKern-Stewart (1957), WMB sample

For the McKern-Stewart method, bias and inaccuracy could only be calculated for the closed ranges provided by scores 1-13. This greatly reduced the number of females that were able to be used in this analysis. In absolute terms, negative bias was more extreme for females than for males though this method underaged both sexes, and inaccuracy was greater for females than for males (Table 8). This is consistent with the stated application of this method.

Table 7: Descriptive statistics for McKern-Stewart scores 1-13

Group Statistics					
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Bias	Female	9	-16.222	13.6613	4.5538
	Male	65	-14.362	10.2620	1.2728
Inaccuracy	Female	9	16.333	13.5116	4.5039
	Male	65	15.227	9.1790	1.1474

Since $p > .05$ for bias according to the Levene's test but $< .05$ for inaccuracy, the t-tests for equality of means were run slightly differently for the two dependent variables in this case so that the resulting figures could be comparable. Ultimately, the means for males and females were not found to be statistically different (Table 2, Appendix V);

however, sample sizes for this test were extremely small, particularly for females, which may have impacted the results of this test.

Gilbert-McKern (1973)

After Brooks (1955) published an analysis of a California Indian series showing female mortality curves far higher than males within the same population, Gilbert and McKern began work on a separate female-specific standard for age estimation. The Gilbert-McKern three-component method was clearly based on the McKern-Stewart method but is recommended for females only (n=108). As with the McKern-Stewart method, the three component scores are added together and the composite score is associated with an age range. Tested ranges and means for each phase are given in Table 10 (males and females) and Table 11 (females only) with the original ranges and means provided by Gilbert and McKern. When both males and females are sampled together, 50.5% of individuals were placed in ranges appropriate to their actual ages. For females alone, 29.6% were placed in appropriate ranges. Since all Gilbert-McKern scores are associated with closed age ranges, average bias and inaccuracy could be calculated for all individuals (n=396). No individuals were given a score of two. For all scores zero through fifteen, the average bias is -3.03 and average inaccuracy is 10.3. For males alone (n=288), these figures are -2.0 and 9.9 respectively; for females alone (n=108), these figures are -5.9 and 11.2 respectively.

Table 8: Gilbert-McKern (1973) tested phase ranges, means, and standard deviations for males and females

Gilbert-McKern (1973) Males and females	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 0 (14-18)	1	0%	25	25	0.0	16.00	2.82
Score 1 (13-24)	1	100%	16	16	0.00	19.80	2.62
Score 2 (16-25)	0	--	--	--	--	--	--
Score 3 (18-25)	5	80.0%	20-38	26.4	6.8	21.50	3.10
Score 4-5 (22-29)	2	50.0%	24-47	35.5	16.3	26.00	2.61
Score 6 (25-36)	2	50.0%	35-54	44.5	13.4	29.62	4.43
Score 7-8 (23-39)	13	46.2%	26-69	43.2	13.6	32.00	4.55
Score 9 (22-40)	12	33.3%	26-71	48.3	13.4	33.00	7.75
Score 10-11 (30-47)	23	39.1%	36-77	52.7	12.7	36.90	4.94
Score 12 (32-52)	17	76.9%	29-59	47.5	6.9	39.00	6.09
Score 13 (44-54)	24	37.5%	33-65	51.7	8.5	47.75	3.59
Score 14-15 (52-59)	296	18.6%	23-99	56.4	13.1	55.71	3.24

Table 9: Gilbert-McKern (1973) tested phase ranges, means, and standard deviations for females only

Gilbert-McKern (1973) Females only	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 0 (14-18)	0	--	--	--	--	--	--
Score 1 (13-24)	0	--	--	--	--	--	--
Score 2 (16-25)	0	--	--	--	--	--	--
Score 3 (18-25)	2	100%	20-24	22.0	2.8	21.50	3.10
Score 4-5 (22-29)	1	0%	47	47	0.0	26.00	2.61
Score 6 (25-36)	1	0%	54	54	0.0	29.62	4.43
Score 7-8 (23-39)	2	50%	32-55	43.5	16.3	32.00	4.55
Score 9 (22-40)	1	0%	60	60	0.0	33.00	7.75
Score 10-11 (30-47)	2	100%	37-38	37.5	0.7	36.90	4.94
Score 12 (32-52)	0	--	--	--	--	--	--
Score 13 (44-54)	3	33.3%	50-58	54.3	16.3	47.75	3.59
Score 14-15 (52-59)	96	27.1%	35-99	60.9	13.7	55.71	3.24

Bias and inaccuracy were calculated for all Gilbert-McKern scores 1-15 and the sample sizes are large; however, the bulk of the tested individuals belonged to the oldest age ranges. In absolute terms, negative bias was more extreme for females than for males, and inaccuracy is greater for females for males (Table 12). This is interesting considering the fact that this method was designed based on an exclusively female sample and is intended for use on females only. One of the potential limitations of this method is the relatively small number of individuals that formed each score of the original sample. With 120 individuals used to create this method, Gilbert and McKern use as few as three individuals to create component descriptions, which may contribute to its tested inaccuracy.

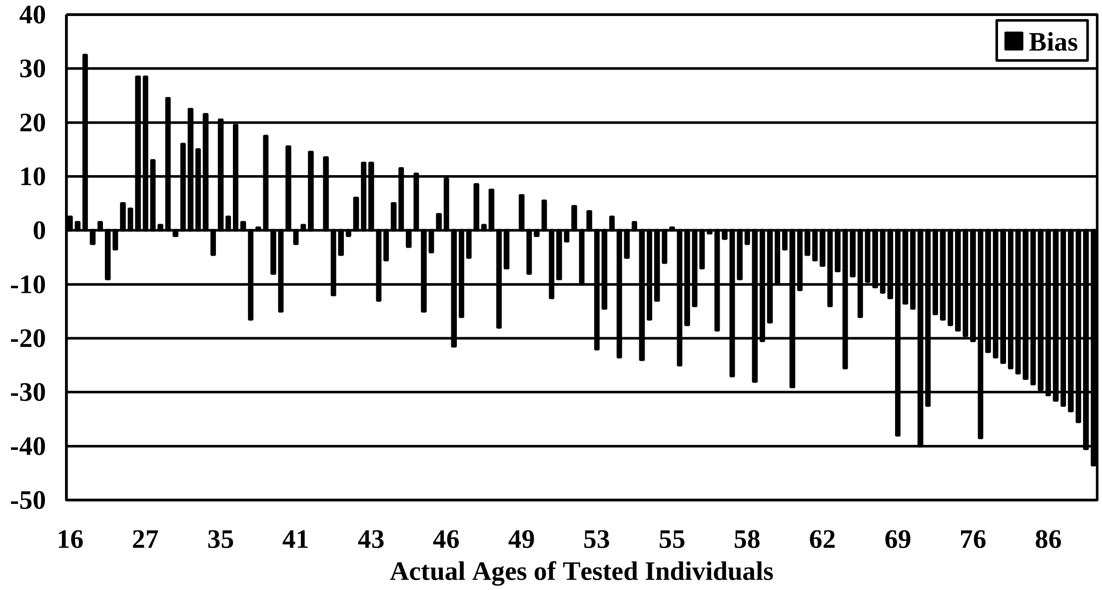


Figure 10: Tested bias and actual age for Gilbert-McKern (1983), WMB sample

Table 10: Descriptive statistics for Gilbert-McKern scores 1-15 for males and females

Group Statistics					
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Bias	Female	108	-5.852	13.4592	1.2951
	Male	288	-1.970	13.0554	.7693
Inaccuracy	Female	108	11.102	9.5557	.9195
	Male	288	9.946	8.6643	.5105

Equal variances were assumed for both bias and inaccuracy based on p-values from the Levene's test. The mean bias was significantly different between male and female categories, which means that this method underages females to a significantly larger extent than males; however, the mean inaccuracy was not significantly different between males and females (Table 3, Appendix V).

<u>Hanihara Suzuki (1978)</u>

The multiple regression and quantification theory model analyses advocated by Hanihara and Suzuki (1978) function similarly to both the McKern-Stewart and Gilbert-McKern methods in the sense that data collection is based on the evaluation of discrete components; however, the Hanihara-Suzuki method dictates that all component scores be entered either into a multiple regression or quantification theory model equation and the resultant figure is then associated with an estimated age. This method is appropriate for both males and females aged 18-38 (n=41; see Figure 8). For the multiple regression analysis, the average bias is 1.18 and the average inaccuracy is -3.68. For the quantification theory model I analysis, these figures are 1.53 and -3.95, respectively.

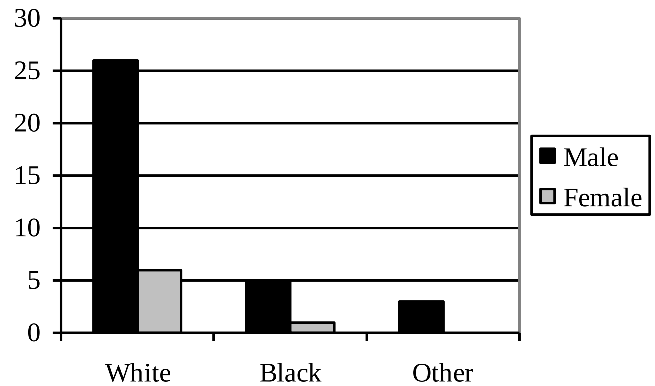


Figure 11: Hanihara-Suzuki sub-sample (n=41)

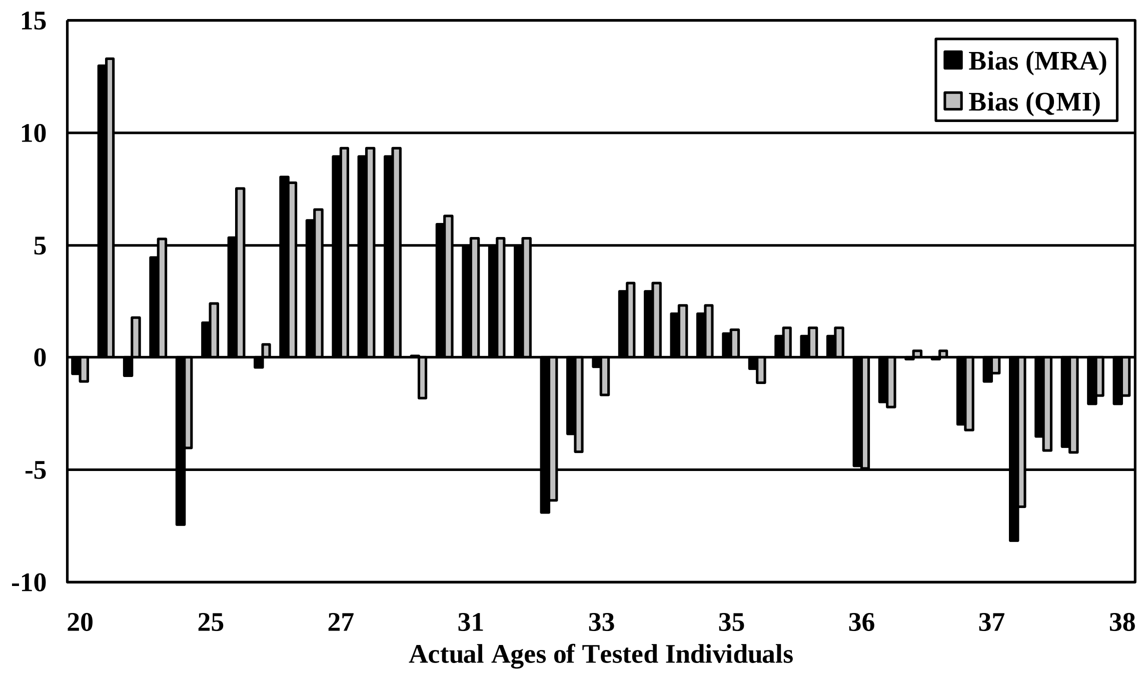


Figure 12: Tested bias and actual age for Hanihara-Suzuki (1978), WMB sample

Though the WMB sample was not racially diverse enough to support testing race as an independent variable contributing to method accuracy, it is interesting to compare the standard error reported by Hanihara and Suzuki (1978) and the standard error of the WMB collection data (see Table 14). The differences between reported and tested standard error are significant for age groups one and two, but these data should be taken with a grain of salt due to the exceedingly small sample sizes. The difference between reported and tested standard error are higher than expected but not significant for age group three, which is worthy of remark because 78% of the WMB sub-sample was Caucasian and this method was developed on an all-Japanese collection. Pal and Tamankar (1983) reported similar results finding that the Hanihara-Suzuki method overestimated age for groups one and two but underaged individuals in age group three.

Though this method is recommended for both males and females, it is impossible to evaluate sex as a contributing variable to error since less than 10% of an already small sub-sample was female in this case. Finally, the difference in standard error between the MRA and QMI methods was not significant, suggesting that they may be used interchangeably.

Table 11: Hanihara-Suzuki (1978) observed results for males and females aged 18-38

Hanihara-Suzuki (1978)	**Number of Individuals**	**Original SE (MRA)**	**Tested SE (MRA)**	**Original SE (QMI)**	**Tested SE (QMI)**
Age group 1 (18-25)	7	2.1960	6.247	1.9055	6.358
Age group 2 (26-30)	8	2.1791	6.836	1.9109	7.045
Age group 3 (31-38)	26	2.7408	3.393	2.7832	3.340

Suchey-Brooks (1990)

Katz and Suchey used Todd's ten-phase system as a foundation for their statistically-based six-phase system, which they developed using 739 male and 486 female individuals from a sample from the Department of Coroner, County of Los Angeles. Though the age range of the sample is as broad as the WMB collection, the average age of the collection is significantly younger than the WMB collection. Katz and Suchey suggested that a phase system would be preferable to a more complicated component system because the components were not found to vary independently, but their analysis suggested that only six phases could be consistently and accurately distinguished (Katz and Suchey 1986). They further simplified the application of this method by providing plaster casts in addition to printed images for each of the phases to be used while assigning phase designations to unknown individuals. Separate casts, images, and ranges are provided for males and females based on separate sampling and analysis.

For the Suchey-Brooks method for males, average bias for all phases one through six is 1.1 and the average inaccuracy is 10.1 for a combined male and female sample (n=396). Tested ranges and averages are presented in Table 15 (males and females) and Table 16 (males only) with the original ranges and averages given by Katz and Suchey. For all individuals tested with this method, coverage was 91.9%. For females alone using the male method (n=108), average bias is 1.4 and average inaccuracy is 10.9. For males (n=288), these figures are -2.0 and 9.8, respectively, and 92.4% of males were placed in appropriate age ranges using this method.

Table 12: Suchey-Brooks (1990) M tested phase ranges, means, and standard deviations for males and females

Suchey-Brooks (1990) M Males and females	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 1 (15-23)	4	50%	16-38	24.75	9.6	18.9	2.3
Score 2 (19-35)	4	75%	24-47	30	11.3	24.7	4.3
Score 3 (22-43)	13	30.8%	25-59	44.1	11.9	28.8	5.9
Score 4 (23-59)	52	94%	23-71	47.1	9.7	36.8	9.6
Score 5 (28-78)	89	97.8%	29-91	51.3	11.4	51.0	13.6
Score 6 (36-87)	234	93.6%	26-99	58.3	13.1	62.7	12.4

Table 13: Suchey-Brooks (1990) M tested phase ranges, means, and standard deviations for males only

Suchey-Brooks (1990) M Males only	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 1 (15-23)	3	33.3%	16-38	26.3	11.1	18.9	2.3
Score 2 (19-35)	2	100%	24-25	24.5	0.7	24.7	4.3
Score 3 (22-43)	9	55.6%	25-59	42.1	13.1	28.8	5.9
Score 4 (23-59)	46	93.5%	23-71	46.7	9.9	36.8	9.6
Score 5 (28-78)	81	98.8%	31-91	50.5	10.8	51.0	13.6
Score 6 (36-87)	147	91.8%	26-96	56.5	12.4	62.7	12.4

For the Suchey-Brooks male-specific method, average bias was positive for females and negative for males, meaning that this method underaged females and overaged males, on average. Average inaccuracy was greater for females than for males in absolute terms (Table 17), which is consistent with the intended application of the method. A Levene's test indicated that the two variances were not significantly different for bias, but they were significantly different for inaccuracy. This test determined the appropriate types of t-tests used to test equality of means.

Table 14: Descriptive statistics for Suchey-Brooks male-specific phases 1-6 for males and females

Group Statistics					
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Bias	Female	108	-1.412	13.8586	1.3335
	Male	288	1.969	12.2416	.7213
Inaccuracy	Female	108	10.921	8.5839	.8260
	Male	288	9.795	7.5807	.4467

The results of the t-tests suggested that this application of the Suchey-Brooks male-specific method did not result in significantly different inaccuracies for males and females, but the difference in bias was significant (Table 4, Appendix V). The trend for overaging younger individuals and underaging older individuals is illustrated in Figure 12. The high adherence of actual age to estimated age ranges is likely attributable to the breadth of the age ranges themselves.

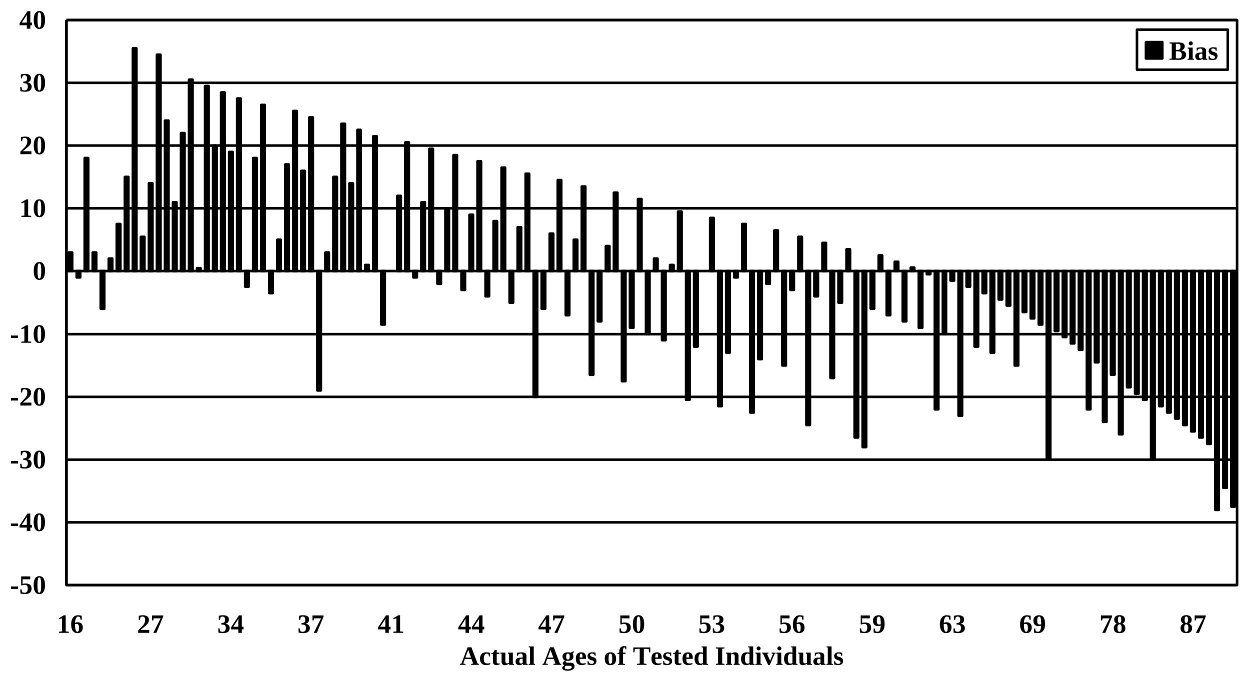


Figure 13: Tested bias and actual age for Suchey-Brooks male-specific (1990) WMB sample

Tested ranges and averages for the Suchey-Brooks female-specific method are presented in Table 19 (males and females) and Table 20 (females only) with the original ranges and averages given by Katz and Suchey. When applied to a combined male and female sample (n=396), the Suchey-Brooks female-specific method resulted in an average bias of -3.4 and average inaccuracy of 9.9 for all scores one through six, and coverage was 92.9%. For males alone (n=288), the average bias is 4.2 and the average inaccuracy is 9.6. For females alone (n=108), these figures are 1.1 and 10.6, respectively, and 93.5% of individuals were placed in appropriate age ranges.

Table 15: Suchey-Brooks (1990) F tested phase ranges, means, and standard deviations for males and females

Suchey-Brooks (1990) F Males and females	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 1 (15-24)	4	50.0%	16-25	21.5	4.4	19.4	2.3
Score 2 (19-40)	4	50.0%	24-54	37.3	15.6	25.0	4.9
Score 3 (21-53)	14	92.9%	25-59	37.43	9.4	30.7	8.1
Score 4 (26-70)	66	95.5%	23-79	48.3	10.1	38.2	10.9
Score 5 (25-83)	94	98.9%	26-88	50.7	10.4	48.1	14.6
Score 6 (42-87)	214	91.1%	31-99	59.4	12.9	60.0	12.4

Table 16: Suchey-Brooks (1990) F tested phase ranges, means, and standard deviations for females only

Suchey-Brooks (1990) F Females only	n	% in range	WMB Tested Sample			Original Sample	
			Range	Mean	SD	Mean	SD
Score 1 (15-24)	1	100%	20	20.0	0.0	19.4	2.3
Score 2 (19-40)	3	33.3%	24-54	41.7	15.7	25.0	4.9
Score 3 (21-53)	5	80%	32-59	43.0	10.9	30.7	8.1
Score 4 (26-70)	7	100%	37-60	49.0	9.6	38.2	109
Score 5 (25-83)	6	100%	45-58	50.8	5.1	48.1	14.6
Score 6 (42-87)	86	95.3%	35-99	62.4	13.4	60.0	12.4

In absolute terms, average positive bias for males was more extreme than for females, though both males and females were overaged by this method, on average. The trend for younger individuals to be overaged and older individuals to be underaged by this method is illustrated in Figure 13, and the high numbers of individuals placed in appropriate age ranges is therefore due to the breadth of the age ranges themselves. Average inaccuracy was somewhat greater for females than for males (Table 21). For both bias and inaccuracy, $p > .05$ for the Levene's test, which means that the two variances were not significantly different and equal variances could be assumed for an independent samples test.

Table 17: Descriptive statistics for Suchey-Brooks female-specific phases 1-6 for males and females

Group Statistics						
	Sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Bias	Female	108	1.148	12.9225	1.2435
	Male	288	4.184	11.4638	.6755
Inaccuracy	Female	108	10.574	7.4476	.7166
	Male	288	9.618	7.4938	.4416

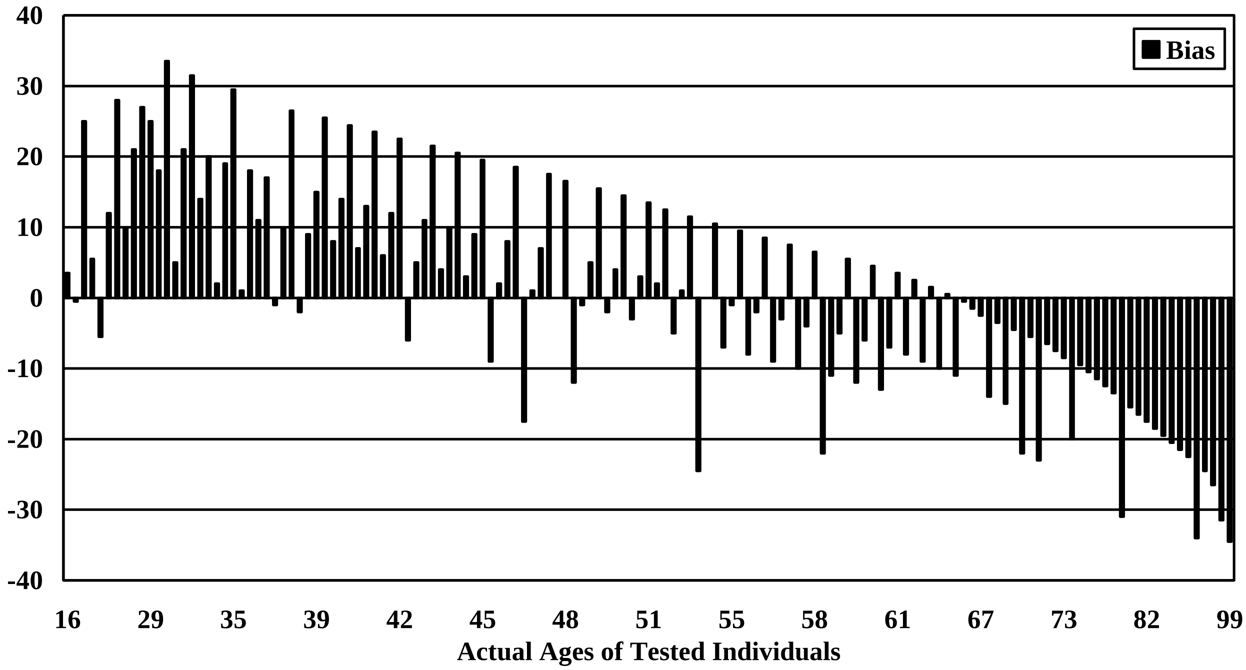


Figure 14: Tested bias and actual age for Suchey-Brooks female-specific (1990), WMB sample

The results of these tests suggested that the Suchey-Brooks female-specific method resulted in no significant difference in mean inaccuracy, but positive bias was significantly greater for females than for males (Table 5, Appendix V).

<u>Berg (2008)</u>

For the Berg female-specific seventh-phase addition to the Suchey-Brooks method, average bias and inaccuracy were not calculated because recommended age ranges for the seven phases were not provided. Instead, only ranges, means, and standard deviations were calculated for each phase using the WMB data (Tables 23; 24).

Table 18: Berg (2008) tested phase ranges, means, and standard deviations for males and females

Berg (2008) Males and females	n	WMB Tested Sample			Original Sample	
		Range	Mean	SD	Mean	SD
Phase 1	4	20	20.0	0.0	19.4	2.3
Phase 2	4	24-54	41.7	15.7	25.0	4.9
Phase 3	13	25-59	37.5	9.8	30.7	8.1
Phase 4	51	23-79	47.8	10.8	38.2	10.9
Phase 5	106	27-88	52.5	10.5	49.7	5.8
Phase 6	101	26-86	53.9	11.5	64.2	9.0
Phase 7	117	31-99	62.1	13.6	74.2	10.9

Table 19: Berg (2008) tested phase ranges, means, and standard deviations for females only

Berg (2008) Females only	n	WMB Tested Sample			Original Sample	
		Range	Mean	SD	Mean	SD
Phase 1	1	20	20.0	0.0	19.4	2.3
Phase 2	3	24-54	41.7	15.7	25.0	4.9
Phase 3	5	32-59	43.0	10.9	30.7	8.1
Phase 4	6	37-59	47.2	9.1	38.2	10.9
Phase 5	12	45-81	57.3	11.0	49.7	5.8
Phase 6	22	35-86	53.5	11.8	64.2	9.0
Phase 7	59	40-99	65.6	12.7	74.2	10.9

Instead of comparing bias and inaccuracy across sex categories to test the utility of the Berg addition, individuals were evaluated using the 68.3% and 95.4% confidence intervals derived from the original standard deviations provided by Berg (2008). Individuals who did not fall within one or two of the published standard deviations were scored as 0 for each category, and individuals who did fall within these ranges were given a score of 1. Means therefore fell between 0 and 1: a mean of 1 indicated perfect adherence to the range, while a mean of 0 would indicate that no individuals fell within the range (Table 25). A t-test for equality of means was used to evaluate whether the difference between these means was significant between males and females (Table 26). The results suggested that this method performed equally well for males and females since the difference between the means proved statistically insignificant.

Table 20: Descriptive statistics for Berg seventh-phase addition

Group Statistics			
	Sex	**N**	**Mean**
Within 1SD	Female	108	.51
	Male	288	.46
Within 2SD	Female	108	.79
	Male	288	.80

Although the Berg (2008) revision was developed using the WMB collection and was intended to improve upon the accuracy of the Suchey-Brooks method for females only, these results are not replicated by this study. For both sexes the number of individuals falling within one or two standard deviations is significantly less than expected: where one standard deviation should include 68.3% of individuals, Berg's one standard deviation included only 51% and 46% of females and males respectively in this test, and where two standard deviations should include 95.4% of individuals, the stated range for two standard deviations is 79% and 80% for females and males respectively (Table 25 and Table 6, Appendix V). This performance is far poorer than expected considering the test was done on the reference sample used by Berg, but could be attributed to higher individual variation than predicted or difficulty in applying a relatively new method.

The second point of interest for the performance of the Berg female-specific method is whether the addition significantly improved the performance of the female-specific Suchey-Brooks method. The purpose of the 7th phase addition was to improve the Suchey-Brooks method's performance in older age categories, which made the WMB collection an ideal sample to develop such a revision. In order to evaluate the Berg method's performance relative to the Suchey-Brooks female-specific method, 1SD and 2SD statistics were also calculated for the Suchey-Brooks female-specific method (Table 27). Suchey-Brooks phases 5-6 were evaluated against Berg phases 5-7, since Berg revised the phase descriptions of phases five and six in addition to adding a seventh phase.

Table 21: Berg and Suchey-Brooks female-specific methods comparison (females only)

Group Statistics			
	Sex	**N**	**Mean**
Within 1SD	SBF	93	.73
	Berg	93	.47
Within 2SD	SBF	93	.98
	Berg	93	.80

Again, despite the fact that the Berg modification used the WMB collection as a reference sample, the performance of the Suchey-Brooks female-specific method for phases five and six was significantly better than the performance of the Berg method for phases 5-7. This is perhaps best explained by the far more constricted ranges of the Berg phases compared to the vast ranges provided for the elderly by Suchey Brooks (1990) rather than any limitations inherent in the method itself, but this will require more testing on additional skeletal collections.

DISCUSSION

The fundamental assumption of this study was that no method would consistently produce the actual age or age range that encompassed the actual range of every individual, and the extent of that error would vary. This assumption proved to be valid, as few individual ages were able to be scored as 'correctly estimated' (i.e., actual age was equal to the midpoint of the scored age range as dictated by the employed method) by any given method within the scope of this study. Sex and ancestry have been suggested as possible variables that can be correlated with error in age estimation, therefore the null hypothesis was that sex and ancestry as independent variables are not statistically significant contributors to overall method performance. Failure to reject the null hypothesis would suggest that no discrete factor could be identified as a leader contributor to error in age estimation except for age itself, and that the individual aging process, i.e. idiosyncrasy and environment, is responsible for variation.

It was initially expected that certain methods would be biased consistent with the results of similar tests on other collections. For example, all methods were expected to perform relatively poorly for middle and advanced ages (40+), and the McKern-Stewart and Gilbert-McKern methods were expected to perform relatively poorly for the opposite sexes because of their specific calibration. It was not expected that ancestry would constitute a statistically significant contributor, as this is a complicated social issue that is least agreed upon in the literature; however, ancestry was not able to be used as a test variable because of the limited racial diversity of the sample.

The most significant outcome of this study was the ultimate failure to reject the null hypothesis based on statistical tests of significance. The difference between bias for males and females was significant for the Gilbert-McKern and both Suchey-Brooks methods, but the difference between inaccuracy for males and females was insignificant for *all methods tested* (see Table 29). It should be noted again that only phases 1-9 for Todd and scores 1-13 for McKern-Stewart were used to establish bias and inaccuracy from range midpoints as closed age ranges were required to make this calculation. Complete samples were used for all other method calculations. The Hanihara-Suzuki method should be considered separately because of its limited range and small sample size, which prevented any comparison between males and females, but the results of this study suggest that both the MRA and QMI are equally successful for estimating age for the 18-38 year old sub-sample. The success of this method in estimating age for the WMB sub-sample is consistent with the results published by Meindl et al (1985), who found the Hanihara-Suzuki method's accuracy to be better than that of other component systems in the 20 to 40-year age range. The slightly preferential reliability of the QMI over the MRA in age groups one and two (18-30 years) reported by Hanihara and Suzuki (1978) was not reproduced in this study.

The poorest performance was predictably the McKern-Stewart method intended for males only, which is easily attributed to the vast differences in sample age distribution between the WMB and Korean War samples. This method consistently underaged both males and females by an average of 15.3 years and was inaccurate by an average of 15.8 years. These figures are similar to the data reported by Klepinger et al (1992) of an average absolute deviation for a combined male and female sample of 15.6. These figures are significantly worse than all other methods tested in this study in addition to being worse than all other aging methods tested by Bedford et al (1993), including auricular surface, femoral radiograph, clavicular radiograph, and summary age (see Table 28).

Table 22: Aging method inaccuracy and bias in years (from Bedford et al 1993)

Skeletal region	Average Inaccuracy	Average Bias
Auricular surface	10.1	1.7
Pubic symphysis	11.0	1.3
Femoral radiograph	12.2	-2.2
Clavicular radiograph	12.8	-3.0
Summary age	8.7	0.0

In terms of average inaccuracy, the performance of the Gilbert-McKern, Suchey-Brooks female-specific, and Suchey-Brooks male-specific methods was not statistically different from one another. This is in contrast to the results published by Klepinger et al (1992) who found that the Gilbert-McKern method performed favorably in terms of absolute deviation from phase means when compared to the Suchey-Brooks method for females. This is not to say that the utility of the Gilbert-McKern method is diminished, as this method's smaller age ranges may be more practical than the larger Suchey-Brooks ranges; however, the coverage of the Suchey-Brooks methods was over 90%, whereas the coverage for the Gilbert-McKern method was 29.6% for females alone and 50.5% for a combined sample. These findings are similar to those published by Suchey (1979), who found that for the Gilbert-McKern method coverage was only 51%. In terms of absolute inaccuracy, the Gilbert-McKern and Suchey-Brooks methods all performed comparably to the data published by Bedford et al (1993) regarding the pubic symphyseal methods and femoral and clavicular radiograph testing. According to their data, summary age methods performed significantly better in terms of accuracy than pubic symphyseal methods, and their auricular surface accuracy may be slightly if not significantly better than the present pubic symphyseal data. While performance in terms of accuracy was different for males and females in absolute terms, it was insignificant in statistical terms for all tested methods.

Table 23: Aging method inaccuracy and bias in years, WMB collection data

Method	Intended for...	Average Bias		Difference Significant?	More accurate for...
		Males	Females		
Todd	Both	-3.904	-2.543	No	Either
McKern-Stewart	Males	-14.362	-16.222	No	Either
Gilbert-McKern	Females	-1.970	-5.852	Yes	Males
Hanihara-Suzuki	Both				
Suchey-Brooks (F)	Females	4.184	1.148	Yes	Females
Suchey-Brooks (M)	Males	-1.969	1.412	Yes	Females

Table 24: Aging method inaccuracy in years, WMB collection data

Method	Intended for...	Average Inaccuracy		Difference Significant?	More accurate for...
		Males	Females		
Todd	Both	8.816	6.478	No	Either
McKern-Stewart	Males	15.227	16.333	No	Either
Gilbert-McKern	Females	9.946	11.102	No	Either
Hanihara-Suzuki	Both				
Suchey-Brooks (F)	Females	9.618	10.574	No	Either
Suchey-Brooks (M)	Males	9.795	10.921	No	Either

In terms of bias, however, the difference between method performance for males and females was significant for the Gilbert-McKern and both Suchey-Brooks methods. The Gilbert-McKern method underaged both sex categories, but interestingly was more negatively biased for females, the intended application of this method. The Suchey-Brooks female-specific method overaged both males and females on average, but overaged females more significantly than males. The Suchey-Brooks male-specific method underaged males and overaged females but did so with the best numbers for bias of all tested methods. Males were underaged more than females were overaged with this method.

The stand-out performer in terms of accuracy was the Todd ten-phase system, though a limited sample was used for the Todd analysis because of the open age range of the terminal phase. The performance of this method in this test was comparable to Bedford et al's (1993) test of summary age methods. Bias and inaccuracy data were not significantly different across sex categories, and although average inaccuracy was much better for the Todd method, it did underage both males and females on average. In addition, this study found that Todd's method overaged individuals younger than 50 but overaged individuals beyond 50, which is consistent with a similar study published by Meindl et al (1983) that reported a very slight tendency to overage in the 20's and 30's, but a far greater tendency to underage in cases over 50. In contrast, Brooks (1955) reported that Todd's method underaged individuals under the age of 30 and overaged individuals beyond 30, but the difference in the age of crossover might be affected by the older average age of the WMB sample. All data sets support the conclusion, however, that this method's non-statistical age range spread is not appropriate for general application.

Conclusions

In terms of performance and applicability for standard contemporary samples it is very apparent that, as expected, methods were limited according to the limitations of the samples upon which they were based. The best predictor for how well each method performed with the WMB collection was how similar in terms of age distribution the WMB sample was to the original sample: for example, the McKern-Stewart method was the most different and subsequently performed the poorest. Sex was not a significant contributor to inaccuracy for any method tested. Bias was statistically different for the Gilbert-McKern and Suchey-Brooks methods, but not consistent with the stated application for each method; for example, the Gilbert-McKern method underaged females to a significantly greater extent than it did males. Although race could not be evaluated as an independent contributor to error in age estimation due to the limitations of the WMB sample, the Hanihara-Suzuki method derived from an all-Japanese sample performed reasonably well on the WMB sample, suggesting that, as expected, race is not a relevant concern in age estimation from the skeleton.

If coverage and accuracy were the single most important factors for selecting an age estimation method the Suchey-Brooks methods should be chosen, though this study suggests that distinguishing between males and females may not be necessary. Though the age ranges for the Suchey-Brooks phases are large, this is inevitable if one needs to ensure with a high reliability that the estimated age range will include the actual age of the individual, meaning that this method might be more useful for a forensic application than a paleodemographic one. The Suchey-Brooks cast system was very useful and eliminated some of the concerns with interpretation of the other methods' illustrations and verbal descriptions. The Todd method performed relatively well for all ages under 50 and is perhaps the simplest to use; however, it may be unwise to assume that an individual is younger than 50 without using a secondary method, thus rendering the use of the Todd method unnecessary. The Hanihara-Suzuki regression methods, though complicated to implement, performed well for the age ranges they were prescribed for; however, again, 18-38 is an extremely limited range and this method should not be used for unknown individuals or populations. The favorable performance of these two methods with the WMB collection may be a valid indicator that further research into this type of method is warranted.

Works Cited

Aiello L.C.; Molleson T. 1993. Are microscopic aging techniques more accurate than macroscopic aging techniques? *Journal of Archaeological Science* 20:689-704.

Ashby R.L.; Roberts S.A.; Mughal M.Z.; Adams J.E.; Ward K.A. 2007. The effect of socioeconomic status upon bone geometry and bone mineral density at different skeletal sites in healthy children. *Bone* 40(6):S24-S24.

Aykroyd, R.G.; Lucy, D.; Pollard, A.M.; Roberts, C.A. 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(1):55-70.

Ängel, J.I. 1969. The bases of paleodemography. *American Journal of Physical Anthropology* 30(3):427-438.

Baccino E.; Ubelaker D.; Hayek L.; 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.

Bedford M.E.; Russell K.F.; Lovejoy C.O.; Meindl R.S.; Simpson S.W.; Stuart-Macadam P.L. 1993. Test of the multifactorial aging method using skeletons with known ages-at-death from the Grant collection. *American Journal of Physical Anthropology* 91:287-297.

Berg G.E. 2008. Pubic bone estimation in adult women. *Journal of Forensic Science* 53(3):569-574.

Boldsen J.; Milner G.R.; Konigsberg L.W.; Wood J.W. 2002. Transition analysis: A new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. *Paleodemography: age distribution from skeletal samples.* Cambridge: Cambridge University Press. p 73-106.

Brooks S.T. 1955. Skeletal age at death: The reliability of cranial and pubic age indicators. *American Journal of Physical Anthropology* 13(4):567-597.

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

Centers for Disease Control and Prevention (CDC). 2008. Deaths: preliminary data for 2006. *National Vital Statistics Reports* 56(16):1-52.

Elliot J.R.; Gilchrist N.L.; Wells J.E. 1996. The effect of socioeconomic status on bone density in a male Caucasian population. *Bone* 18(4):371-373.

Galera V.; Ubelaker D.; Hayek L. 1995. Comparison of macroscopic cranial estimations of age estimation applied to skeletons from the Terry collection. *Journal of Forensic Science* 43:933-939.

Gilbert, B.M. 1973. Misapplication to females of the standard for aging the male *os pubis*. *American Journal of Physical Anthropology* 38:39-40.

Gilbert B.M. and McKern T.W. 1973. A method for aging the female *os pubis*. *American Journal of Physical Anthropology* 19:237-244.

Gillett R.M. 1991. Determination of age at death in human skeletal remains: A comparison of two techniques. *Journal of Forensic Science* 6(2):179-189.

Hanihara K. 1952. Age change in the male Japanese pubic bone. *Journal of the Anthropological Society of Nippon* 62:245-260.

Hanihara K. and Suzuki T. 1978. Estimation of age from the pubic symphysis by means of multiple regression analysis. *American Journal of Physical Anthropology* 48:233-240.

Hermann B. and Bergfelder T. 1977 Über den diagnostischen Wert des sogenannten Geburtstrauma am Schambein bei der Identifikation. *Zeischrift für Rechtsmedizin* 81:73-78.

Heyman J. and Lundquist A. 1932. The symphysis pubis in pregnancy and parturition. *Acta Obstetrica et Gynecologica Scandinavica* 12:191-223.

Hoppa R.D. 2000. Population variation in osteological aging criteria: An example from the pubic symphysis. *American Journal of Physical Anthropology* 111:185-191.

Katz D. and Suchey J.M. 1986. Age estimation of the male *os pubis*. *American Journal of Physical Anthropology* 69:427-435.

Katz D. and Suchey J.M. 1989. Race differences in pubic symphyseal aging patterns in the male. *American Journal of Physical Anthropology* 80:167-172.

Kemkes-Grottenthaler A. 1996. Critical evaluation of osteomorphognostic methods to estimate adult age at death: A test of the "complex method". *Homo* 46:280-292.

Kimmerle E.H.; Konigsberg L.W.; Jantz R.L.; Baraybar J.P. 2008a. Analysis of age-at-death estimation through the use of pubic symphyseal data. *Journal of Forensic Science*53(3):558-565.

Kimmerle E.H.; Prince D.A.; Berg G.E. 2008b. Inter-observer variation in methodologies involving the pubic symphysis, sternal rib ends, and teeth. Journal of Forensic Science 3(594):600.

Klepinger L.L.; Katz D.; Micozzi M.S.; Carroll L. 1992. Evaluation of cast methods for estimating age from the *os pubis*. *Journal of Forensic Science* 37:763-770.

Konigsberg, L.W.; Frankenberg, S.R. 1994. Paleodemography: “Not quite dead” 3:92-105.

Konigsberg, L.W.; Hermann, N.P.; Wescott, D.J.; Kimmerle, E.H. 2008. Estimation and evidence in forensic anthropology: age-at-death. *Journal of Forensic Science* 53(3):541-557

Komar D. 2003. Lessons from Srebrenica: The contributions and limitations of physical anthropology in identifying victims of war crimes. *Journal of Forensic Science* 48(4):713-716.

Krogman W.M. 1962. *The human skeleton in forensic medicine.* Springfield, IL: Charles C. Thomas.

Lovejoy C.O.; Meindl R.S.; Mensforth R.P.; Barton T.J. 1985. Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. *American Journal of Physical Anthropology* 68:1-14.

Lucy, D.; Aykroyd, R.G.; Pollard, A.M.; Solheim, T. A Bayesian approach to adult human age estimation from dental observation by Johanson’s age changes. *Journal of Forensic Science* 41(2): 189-194.

McKern, T.W. and Stewart T.D. 1957. Skeletal age changes in young American males, analyzed from the standpoint of age identification. Natick, MA: Headquarters Quartermaster Research and Development Command.

Meindl R.S.; Lovejoy C.O.; Mensforth R.P. 1983. Skeletal age at death: Accuracy of determination and implications for human demography. *Human Biology* 55:73-87.

Meindl R.S.; Lovejoy C.O.; Mensforth R.P.; Walker R.A. 1985. A revised method of age determination using the *os pubis*, with a review of tests of accuracy of other current methods of pubic symphyseal aging. *American Journal of Physical Anthropology* 68:29-45.

Owings Webb, P. A. and Suchey J.M. 1985. Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *American Journal of Physical Anthropology* 68:457-466.

Pal, G.P. and Tamankar, B.P. 1983. Determination of age from pubic symphysis. *Indian Journal of Medical Research* 99:694-701.

Putschar, W.G. 1931. Entwicklung, wachstum und pathologie der beckenverbindungen des menschen. Gustav Fischer, Jena.

Putschar, W.G. 1976. The structure of the human symphysis pubis with special consideration of parturition and its sequelae. *American Journal of Physical Anthropology* 45(3):589-594.

Rissech C.; Schmitt A.; Malagosa A; Cunha E. 2004. Influencia de las patologías en los indicadores de edad adulta del coxal: estudio preliminar. *Antropologia Portuguesa* 21:265-277.

Rissech C.; Estabrook G.F.; Malgosa A.; Cunha E. 2006. Using the acetabulum to estimate age at death of adult males. *Journal of Forensic Science* 51:213-229.

Rogers T. and Saunders S.R. 1994. Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences* 39(4):1047-1056.

Sakaue K. 2006. Application of the Suchey-Brooks system of pubic age estimation to recent Japanese skeletal material. *Anthropological Science* 114(1):59-64.

Saunders S.R.; Fitzgerald C.; Rogers T.; Dudar C.; McKillop H. 1992. A test of several methods of skeletal age estimation using a documented archaeological sample. *Canadian Society of Forensic Science Journal* 24:97-118.

Scheuer L. and Black S. 2000. Developmental juvenile osteology. New York: Academic Press.

Sheldon W.H. 1940. *The varieties of human physique, an introduction to constitutional psychology*. New York.

Sinha A. and Gupta V. 1995. A study on estimation of age from pubic symphysis. *Forensic Science International* 1(73):78.

Stewart 1957. Distortion of the pubic symphyseal surface in females and its effect on age determination. *American Journal of Physical Anthropology* 15: 9-18.

Suchey J.M. 1979. Problems in the aging of females using the *os pubis*. *American Journal of Physical Anthropology* 51(3):467-470.

Suchey J.M. and Katz D. 1986. Skeletal age standards derived from an extensive multi-racial sample of modern Americans. Paper presented at the 55th annual meeting of the American Association of Physical Anthropologists, Albuquerque, NM.

Suchey J.M.; Brooks S.T.; Katz D. 1988. Instructions for use of the Suchey-Brooks system for age determination of the female *os pubis.* Instructional materials accompanying female pubic symphyseal models of the Suchey-Brooks system. Distributed by France Casting, Colorado.

Todd T.W. 1920. Age changes in the pubic bone: I. The male White pubis. *American Journal of Physical Anthropology* 3:285-334.

Todd T.W. 1921. Age changes in the pubic bone: II. The pubis of the male Negro-White hybrid: III. The pubis of the White female: IV. The pubis of the female Negro-White hybrid. *American Journal of Physical Anthropology 4*:1-70.

Todd T.W. 1923. The pubic symphysis of the guinea-pig in relation to pregnancy and parturition. *American Journal of Anatomy* 31(4):345-357.

Ubelaker D.H. 2008. Issues in the global applications of methodology in forensic anthropology. *Journal of Forensic Science* 53(3):606-607.

Walde, J. 1962. Obstetrical and gynæcological back and pelvic pains, especially those contracted during pregnancy. *Acta Obstetrica et Gynecologica Scandinavica* 41(4):11-53

Washburn S.L. 1948. Sex differences in the pubic bone. *American Journal of Physical Anthropology* 6:199-207.

APPENDIX I

Todd (1920)

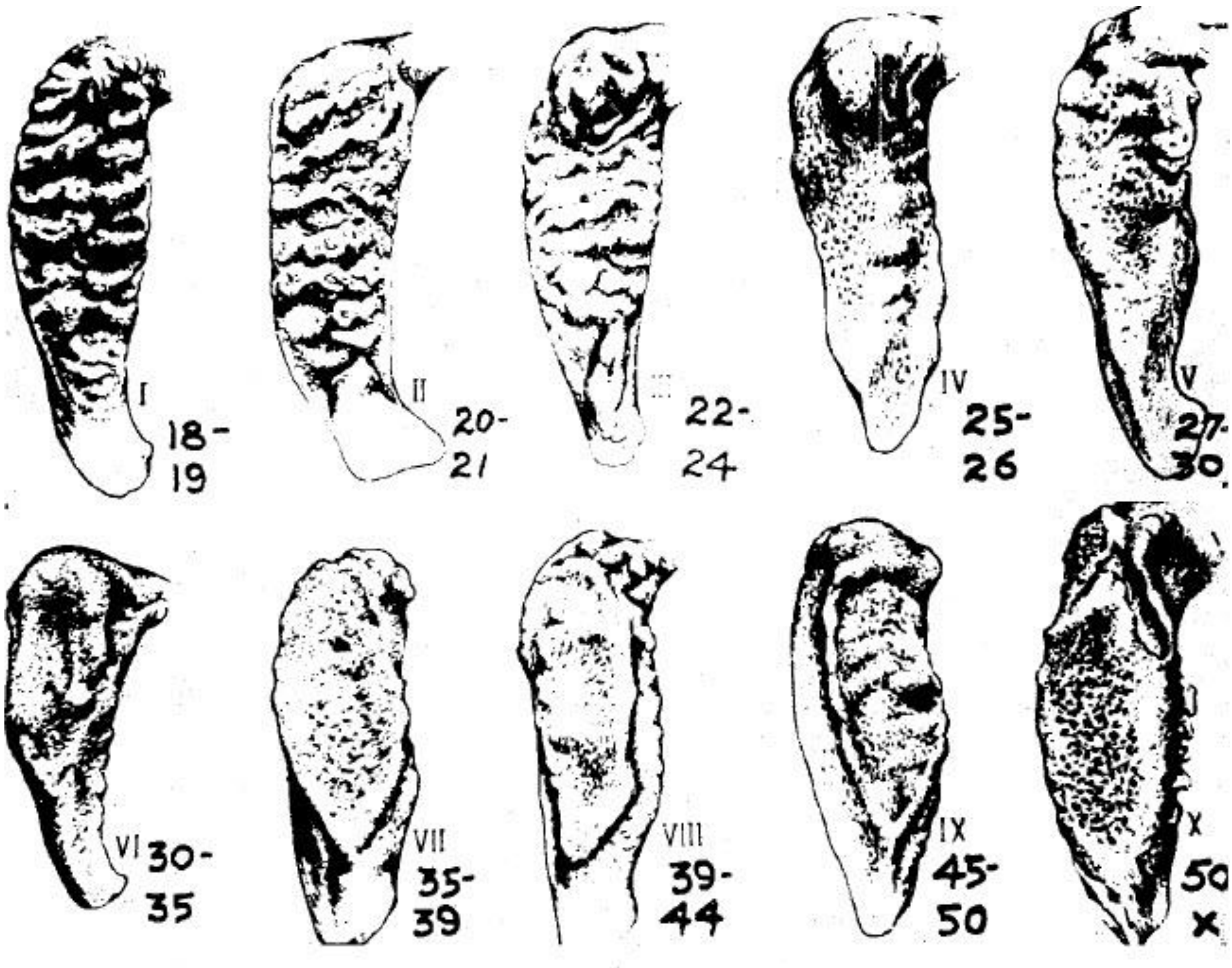


Figure 15: Modal Standards of Todd's 10 Typical Phases (after G. Neumann *in* Stewart and McKern 1957)

<u>**McKern-Stewart (1957)**</u>

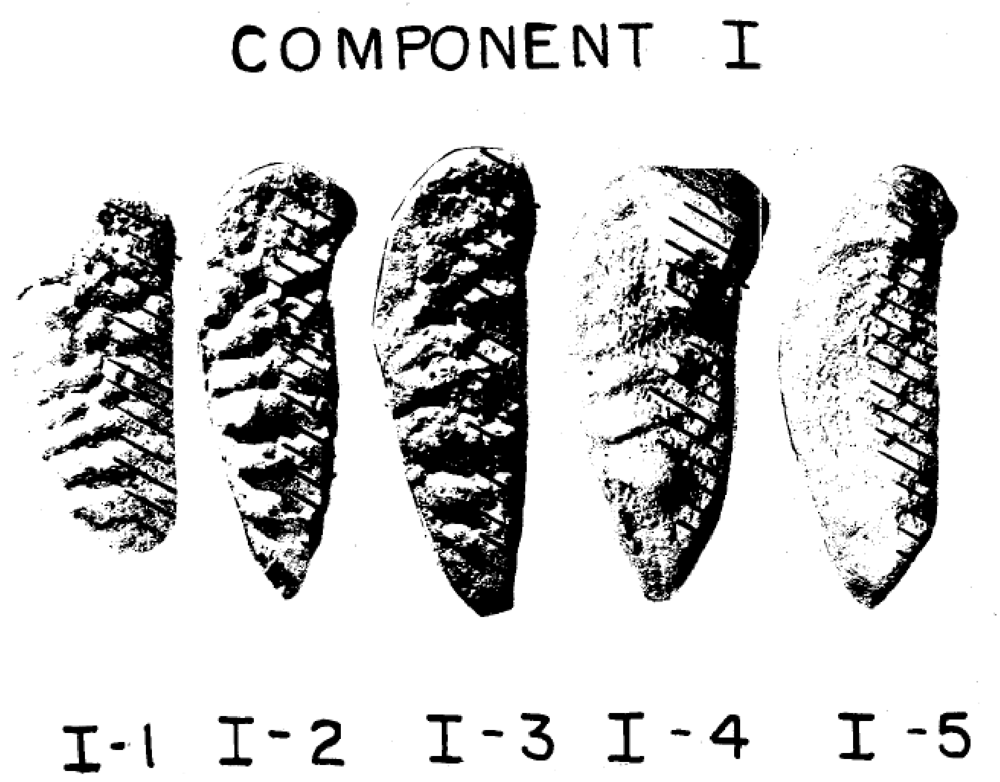


Figure 16: McKern-Stewart Component 1 (Stewart and McKern 1957)

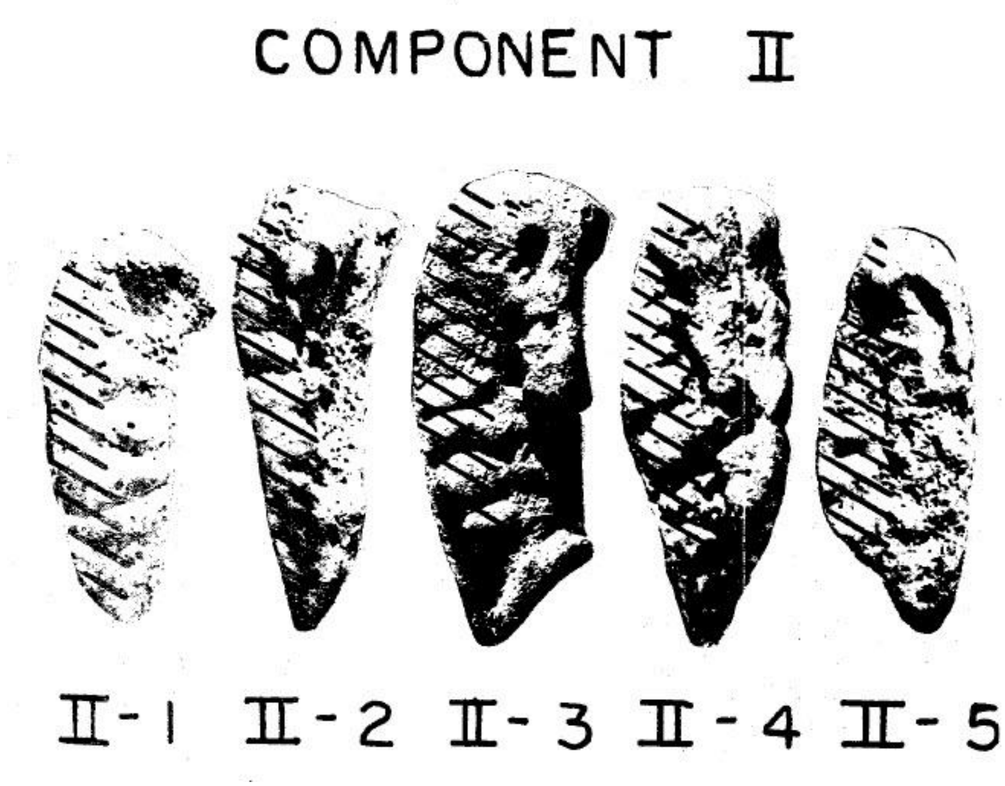


Figure 17: McKern-Stewart Component 2 (Stewart and McKern 1957)

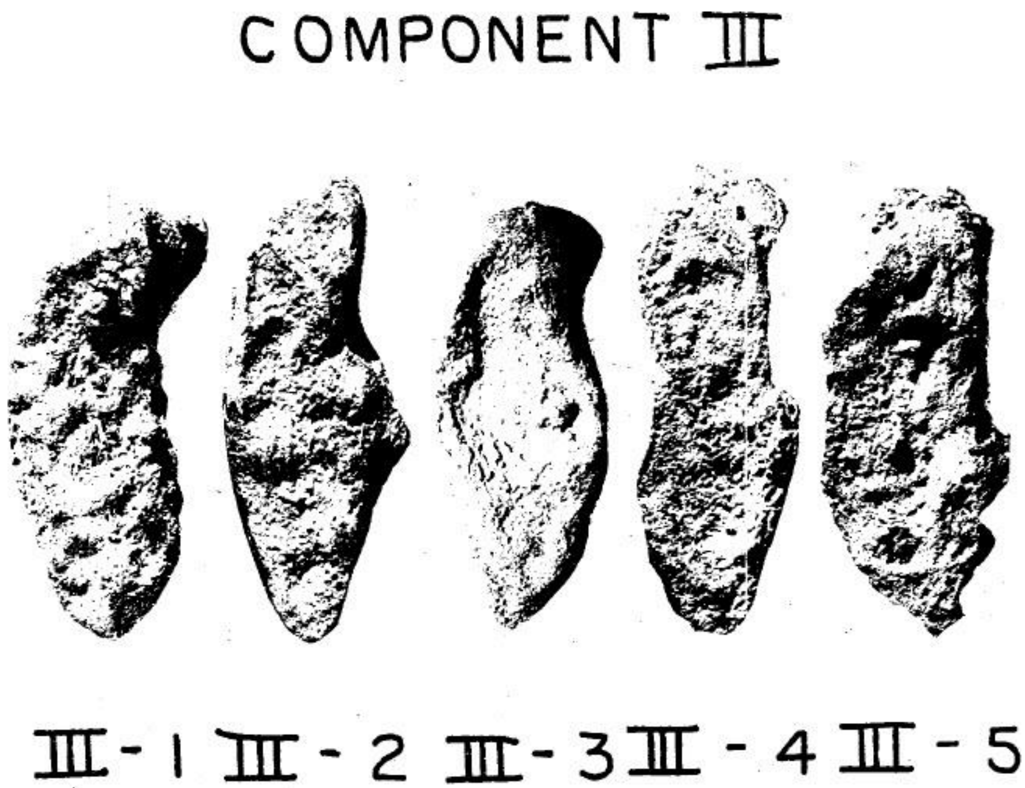


Figure 18: McKern-Stewart Component 3 (Stewart and McKern 1957)

Gilbert-McKern (1973)

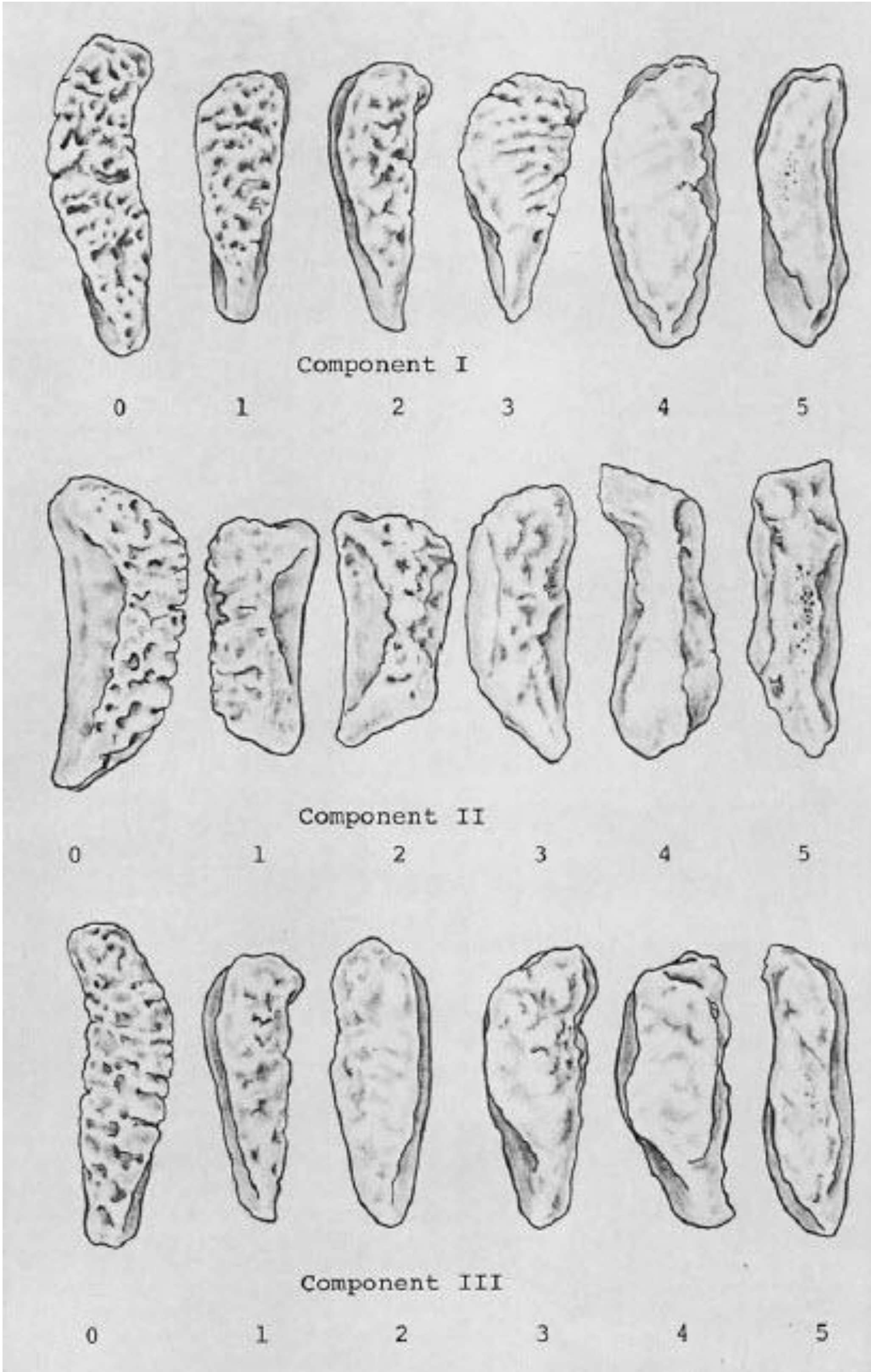


Figure 19: Gilbert-McKern Components 1, 2, 3 (Gilbert and McKern 1973)

Hanihara-Suzuki (1978)

Figure 20: Hanihara-Suzuki illustration of four stages of pubic age (Hanihara and Suzuki 1978)

Berg (2008)

Figure 2: Line drawings of each phase. Top row: phase V from the Balkan sample (left), phase V from the WMB sample (right). Middle row: phase VI from the Balkan sample (left), phase VI from the WMB sample (right). Bottom row: phase VII from the Balkan sample (left), phase VII from the WMB sample (right) (Berg 2008

Appendix II

Todd (1920)

Case Number

Phase (I-X)

Phase Descriptions

I: Typical adolescent ridge and furrow formation with no sign of margins and no ventral beveling.

II: Foreshadowing of ventral beveling with slight indication of dorsal margin.

III: Progressive obliteration of ridge and furrow system with increasing definition of dorsal margin and commencement of ventral rarefaction (beveling).

IV: Completion of definite dorsal margin, rapid increase of ventral rarefaction and commencing delimitation of lower extremity.

V: Commencing formation of upper extremity with increasing definition of lower extremity and possible sporadic attempts at formation of ventral rampart.

VI: Development and practical completion of ventral rampart with increasing definition of extremities.

VII: Changes in symphyseal face and ventral aspect of pubis consequent upon diminishing activity, accompanied by bony outgrowths into pelvic attachments of tendons and ligaments.

VIII: Smoothness and inactivity of symphyseal face and ventral aspect of pubis. Oval outline and extremities clearly defined but no "rim" formation or lipping.

IX: Development of "rim" on symphyseal face with lipping of dorsal and ventral margins.

X: Erosion of and erratic, possibly pathological, osteophytic growth on symphyseal face with breaking down of ventral margins.

McKern-Stewart (1957)

Case Number	Component Scores (0-5)
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________
______________	______________

Component Descriptions

Component I

0: Dorsal margin absent

1: Slight margin formation first appears in middle third of the dorsal border

2: Dorsal margin extends along entire dorsal border

3: Filling in of grooves and resorption of ridges to form a beginning plateau in middle third of dorsal demiface

4: Plateau extends over most of dorsal demiface with vestiges of billowing

5: Billowing disappears completely; surface of demiface becomes flat and granulated

Component II

0: Ventral beveling absent

1: Ventral beveling present only at superior extremity of border

2: Bevel extends inferiorly along ventral border

3: Ventral rampart begins by means of bony extensions from either or both extremities

4: Rampart is extensive but gaps evident on earlier ventral border, especially upper two-thirds

5: Rampart complete

Component III

0: Symphyseal rim absent

1: Partial dorsal rim present; round and smooth, elevated above surface

2: Dorsal rim complete; ventral rim beginning to form

3: Symphyseal rim complete; surface is finely-grained, irregular and undulating

4: Rim breakdown, rim is sharp rather than round, with lipping at ventral edge

5: Further breakdown of rim and rarefaction of symphyseal face; erratic ossification along ventral rim

Gilbert-McKern (1973)

Case Number | Component Scores (0-5) | Component Descriptions

Component I

0: Ridges and furrows distinct and billowed; dorsal margin undefined

1: Ridges begin to flatten; dorsal margin begins in mid-third of demiface

2: Dorsal demiface spreads ventrally, becomes wider; dorsal margin extends

3: Dorsal demiface smooth; margin may be narrow or indistinct from face

4: Demiface complete and unbroken; broad and fine-grained; may exhibit vestigial billowing

Component II

0: Ridges and furrows distinct

1: Furrows begin to fill in beginning inferiorly forming an expanding rampart; lateral edge of which is a distinct curved line

2: Continued fill in of furrows and expansion of demiface from both superior and inferior ends; rampart spreads laterally along ventral edge

3: All but one-third of ventral demiface filled in with fine-grained bone

4: Ventral rampart presents broad, complete, fine-grained surface from the pubic crest to inferior ramus

5: Ventral rampart begins to break down; pitted appearance

Component III

0: Symphyseal rim absent

1: Rim begins in mid-third of dorsal surface

2: Dorsal rim complete

3: Rim extends from superior and inferior ends of symphysis until all but one-third of the ventral aspect complete

4: Symphyseal rim complete

5: Ventral margin of dorsal demiface may break down or round off

Hanihara-Suzuki (1978)

Case Number | Variable Scores (1-4) | Feature Descriptions

Horizontal ridges and furrows (X_1)

1: Distinct

2: Furrows become shallow

3: Trace

4: No longer visible

Pubic tubercle (X_2)

1: Attached by cartilage

2: United

Lower end (X_3)

1: Indistinct

2: Narrow ridge

3: Broad ridge

Dorsal margin (X_4)

1: None

2: Interrupted narrow ridge

3: Narrow ridge over full length

4: Broad ridge

Superior ossific nodule (X_5)

1: None

2: Present

3: No longer visible

Ventral beveling (X_6)

1: None

2: Incomplete

3: Completed over full length

4: Upper part no longer visible

Symphyseal rim (X_7)

1: Incomplete

2: Whole symphyseal rim bordered by broad rim

Suchey-Brooks (1990)

Case Number	Phases (1-6)

Phase Descriptions

1: Symphyseal face has billowing surface usually extending to pubic tubercle. Horizontal ridges well-marked and ventral beveling may be commencing. Ossific nodules may occur on upper extremity, but *no delimitation of either extremity.*

2: Symphyseal face may show ridge development. *Face has commencing delimitation of lower and/or upper extremities with or without ossific nodules*. Ventral rampart may be beginning as an extension of either or both extremities.

3: Symphyseal face shows lower extremity and *ventral rampart in process of completion*. May be continuation of fusing ossific nodules forming upper extremity and along ventral border. Symphyseal face is smooth or ridged. Dorsal plateau complete. No lipping of dorsal margin; no bony outgrowths.

4: Symphyseal face fine grained; may have remnants of ridges. *Oval outline usually complete; hiatus at ventral rim may occur*. Pubic tubercle fully separate. Bony outgrowths ventrally on inferior portion; slight lipping may occur on dorsal border.

5: *Symphyseal face completely rimmed with slight depression of the face itself.* Moderate lipping on dorsal border; more outgrowths on ventral border. Little or no rim erosion; breakdown may occur on superior ventral border.

6: *Symphyseal face may show ongoing depression as rim erodes*. Ventral ligamentous attachments marked; pubic tubercle often appears as separate bony knob. Face may be pitted or porous; shape may be irregular with crenulations.

Berg (2008)

Case Number

Phases (I-IV)

Phase Descriptions

I: See Suchey-Brooks (1990)

II: See Suchey-Brooks (1990)

III: See Suchey-Brooks (1990)

IV: See Suchey-Brooks (1990)

V: The rim is complete, but the symphyseal face may show a slight depression as it begins to erode. The pubic tubercle is separate from the face. Bone is compact, though a slight amount of porosity may be present (less than 15% of the face). Decision-making traits: *if the articular surface has less than 15% porosity anywhere on surface, and osteoporosity is nearly absent.*

VI: The symphyseal face is usually depressed and the rim begins to erode beginning with the superior ventral aspect. Quality of bone is breaking down; symphyseal face is eroded (porosities or channels). Osteoporosis is mild to moderate; lipping can be present. Decision-making traits: *if less than 50% of the symphyseal face is porous and lipping is mild to moderate.* If the symphyseal face is borderline, use osteoporosity: if moderate to severe, score as VII.

VII: The symphyseal face is extremely porous and eroded with >50% of its surface. Osteoporosity is present and is typically moderate to severe. The symphyseal face appears to be relatively flat, since the rim is highly eroded and is losing definition. The ventral surface of the symphysis is typically scarred or has striated bone with ligamentous outgrowths near the obturator foramen. Lipping of the articular surface is often moderate, but may be mild or severe. This character is highly variable.

APPENDIX III

		n*	Age range	Midpoint	Mean	SD
Todd (1920)	I	3	18-19	18.5	18	0.00
	II	2	20-21	20.5	20.50	0.71
	III	14	22-27	24.5	23.71	1.68
	IV	5	25-26	25.5	25.60	0.55
	V	15	27-30	28.5	29.33	2.74
	VI	18	30-35	32.5	33.56	2.15
	VII	22	35-39	37	36.91	2.83
	VIII	29	39-44	41.5	41.07	4.97
	IX	66	45-50	47.5	52.62	9.29
	X	48	50+	--	63.56	11.39
McKern-Stewart (1957)	0	7	>17	--	17.29	0.49
	1-2	76	17-20	18.5	19.04	0.79
	3	43	18-21	19.5	19.79	0.85
	4-5	51	18-23	20.5	20.84	1.13
	6-7	26	20-24	22	22.42	0.99
	8-9	36	22-28	25	24.14	1.93
	10	19	23-28	25.5	26.05	1.87
	11-13	56	23-39	31	29.18	3.33
	14	31	29+	--	35.84	3.89
	15	4	36+	--	41.00	6.22
Gilbert-McKern (1973)	0	2	14-18	16	16.00	2.82
	1	12	13-24	18.5	19.80	2.62
	2	13	16-25	20.5	20.15	2.19
	3	4	18-25	21.5	21.50	3.10
	4-5	7	22-29	25.5	26.00	2.61
	6	8	25-36	30.5	29.62	4.43
	7-8	14	23-39	31	32.00	4.55
	9	5	22-40	31	33.00	7.75
	10-11	11	30-47	38.5	36.90	4.94
corrected	12	12	32-52	42	39.00	6.09
	13	8	44-54	49	47.75	3.59
	14-15	7	52-59	55.5	55.71	3.24
Suchey-Brooks (1986)	1	121	15-23	19	18.9	2.30
	2	81	19-35	27	24.7	4.30
	3	43	22-43	32.5	28.8	5.90
	4	153	23-59	41	36.8	9.60
Male	5	241	28-78	53	51.0	13.6
	6	100	36-87	61.5	62.7	12.4
Suchey-Brooks (1990)	1	48	15-24	19.5	19.4	2.6
	2	47	19-40	29.5	25.0	4.9
	3	44	21-53	37	30.7	8.1
	4	39	26-70	48	38.2	10.9
Female	5	44	25-83	54	48.1	14.6
	6	51	42-87	64.5	60.0	12.4
Berg (2008) Female§	5	18	--	--	49.7	5.8
	6	27	--	--	64.2	9.0
	7	50	--	--	74.2	10.9

* Phases given as a range in the original publication (e.g. 6-7) were excluded from this analysis
§ Original data utilized both the WMB collection and a Balkan sample; original WMB data are used here

<u>APPENDIX IV</u>

		Original	LA Coroner	USA (total)*
Todd (1920)	I	0.00	1.46	-
	II	0.71	1.74	-
	III	1.68	1.90	-
	IV	0.55	3.72	-
	V	2.74	4.58	-
	VI	2.15	5.89	-
	VII	2.83	9.47	-
	VIII	4.97	9.43	-
	IX	9.29	13.64	-
	X	11.39	12.40	-
McKern-Stewart (1957)	0	0.49	3.54	-
	1	0.79	2.14	-
	2	0.79	0.58	-
	3	0.85	1.98	-
	4	1.13	0.82	-
	5	1.13	1.95	-
	6	0.99	2.13	-
	7	0.99	2.30	-
	8	1.93	4.56	-
	9	1.93	-	-
	10	1.87	6.52	-
	11	1.87	12.48	-
	12	1.87	9.94	-
	13	3.33	9.96	-
	14	3.89	12.16	-
	15	6.22	13.89	-
Suchey-Brooks (1986) male	1	2.30	-	3.46
	2	4.30	-	8.36
	3	5.90	-	9.77
	4	9.60	-	12.73
	5	13.6	-	15.14
	6	12.4	-	14.36
Suchey-Brooks (1990) female	1	2.6	-	4.44
	2	4.9	-	10.60
	3	8.1	-	11.74
	4	10.9	-	13.22
	5	14.6	-	18.21
	6	12.4	-	20.62

* includes the Forensic Data Bank, Gilbert and McKern (1973) collection, Korean War Dead (Stewart 1957), Los Angeles Coroner sample, and the Robert J. Terry Anatomical Collection (from Kimmerle et al 2008b)

APPENDIX V

Table 1: Todd phases 1-9: Levene's and t-tests for equality of means across sex categories

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	1.986	.161	.573	146	.567	1.3605	2.3733	-3.3299	6.0509
Inaccuracy	Equal variances assumed	3.514	.063	-1.395	146	.165	-2.3377	1.6756	-5.6492	.9737

Table 2: McKern-Stewart scores 1-13: Levene's and t-tests for equality of means across sex categories

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	2.509	.118	-.489	72	.626	-1.8607	3.8032	-9.4422	5.7208
Inaccuracy	Equal variances not assumed	4.501	.037	.238	9.07	.817	1.1068	4.6477	-9.3952	11.6087

Table 25: Gilbert-McKern scores 1-15: Levene's and t-tests for equality of means across sex categories

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	.636	.426	-2.613	394	.009	-3.8814	1.4856	-6.8021	-.9607
Inaccuracy	Equal variances assumed	2.801	.095	1.149	394	.251	1.1557	1.0059	-.8220	3.1333

Table 4: Suchey-Brooks phases 1-6: Levene's and t-tests for equality of means across sex categories

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	2.918	.088	-2.359	394	.019	-3.308	1.4331	-6.1983	-.5633
Inaccuracy	Equal variances not assumed	7.755	.006	1.199	173	.232	1.1262	.9390	-.7273	2.796

Table 5: Suchey-Brooks female-specific phases 1-6: Levene's and t-test for equality of means across sex categories

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	3.466	.063	2.265	394	.024	3.0359	1.3402	.4010	5.6707
Inaccuracy	Equal variances assumed	.451	.502	1.133	394	.258	.9560	.8441	-.7036	2.6156

Table 26: Berg female-specific phases 1-6: Levene's and t-test for equality of means across sex

Independent Samples Test										
		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Bound	Upper Bound
Bias	Equal variances assumed	.123	.726	-.177	394	.860	-.008	.046	-.098	.082
Inaccuracy	Equal variances assumed	.667	.415	.903	394	.367	.051	.056	-.060	.162