

Linda L. Klepinger,¹ Ph.D.; Darryl Katz,² Ph.D.;
Marc S. Micozzi,³ M.D., Ph.D.; and Laurie Carroll,³ M.S.

Evaluation of Cast Methods for Estimating Age from the *Os Pubis*

REFERENCE: Klepinger, L. L., Katz, D., Micozzi, M. S., and Carroll, L., "Evaluation of Cast Methods for Estimating Age from the *Os Pubis*," *Journal of Forensic Sciences*, JFSCA, Vol. 37, No. 3, May 1992, pp. 763-770.

ABSTRACT: Age-related changes of the pubic symphysis have gained wide acceptance among physical anthropologists as one basis for estimating adult skeletal age. However, these methods have not been tested by independent observers on large samples with known ages at death that represent contemporary populations. In this study 202 female and 116 male pairs of pubes collected at autopsy were blindly evaluated for age using the McKern-Stewart or Gilbert-McKern and Suchey-Brooks methods. Performance of the methods was measured by mean absolute deviation of true age from scored interval means and frequencies of true age falling within ± 1 and ± 2 standard deviations from the mean. The results suggest that the Suchey-Brooks methods are superior in forensic applications and that the racially specific refinement for males should be used. Age estimates should include ± 2 standard deviations, and chances of error should be considered, especially when advanced age or debilitation is suspected.

KEYWORDS: physical anthropology, age determination, musculoskeletal system, *Os pubis*, human identification

Of the various skeletal indicators used by physical anthropologists for estimating the age of older adolescents and adults, no single criterion has received more attention than the morphological development of the pubic symphyseal face. Brooks and Suchey [1] have summarized the historical development of various techniques employed in pubic age assessment. Currently, two of the methods frequently applied by forensic anthropologists in North America are the three component methods devised by McKern and Stewart [2] and Gilbert and McKern [3] for males and females, respectively. However, Suchey and Brooks [4] have recently developed separate six-phase model systems for males and females that suggest potential advantages for forensic anthropologists. A common feature to all of these methods is the availability of model casts for comparison with the specimen, a feature that increases ease of use considerably over photographs. The Suchey-Brooks systems are based on large samples of known ages collected from autopsies performed at the Office of the Chief Medical Examiner-Coroner, County of Los Angeles. These large multiracial samples representing adolescence through extreme

Received for publication 31 July 1991; revised manuscript received 4 Oct. 1991; accepted for publication 8 Oct. 1991.

¹Associate professor, Department of Anthropology, University of Illinois, Urbana, IL.

²Professor, Department of Mathematics, California State University, Fullerton, CA.

³Director and special projects manager, Division of Collections and Research, respectively, National Museum of Health and Medicine, Armed Forces Institute of Pathology, Washington, DC.

old age are more appropriate to the forensic science situation than either the McKern-Stewart or Gilbert-McKern databases.

But does the superiority of the Suchey-Brooks database translate into more successful age assignments to unidentified skeletons? In two previous studies [5,6] a group of forensic anthropologists was asked to assess the age of eleven male pubic bones by the McKern-Stewart method and eleven female pubic bones by the Gilbert-McKern method. The results were somewhat discouraging; the known age fell within the estimated range 76% of the time for males and only 51% of the time for females. Meindl and coworkers [7] reported average errors in age estimation of 16.1 and 8.9 years for males and females, respectively, when applying the McKern-Stewart and Gilbert-McKern systems to a sample of 64 males and 30 females from the Hamann-Todd collection that ranged in age from 20 to 69 years. This study evaluates the performance of the McKern-Stewart, Gilbert-McKern, and Suchey-Brooks methods over a large number of cases. For the males the comparative success of the 1988 Suchey-Brooks method [4] based on pooled racial standards and the 1989 refinement [8] based on separate racial criteria was also evaluated.

Materials and Methods

Two hundred and two paired female pubic bones collected by Judy Suchey from the Office of the Chief Medical Examiner-Coroner, County of Los Angeles were blindly evaluated for age at death by one of us (L.L.K.) using both the Gilbert-McKern [3] and Suchey-Brooks [4] methods. Individuals represented in the sample spanned ages 15 through 95, and all known ages were confirmed by both death and birth certificates. This test sample was not included in the original sample upon which the Suchey-Brooks method was based.

Similarly, 116 male pubic symphyses were collected by Marc Micozzi and Laurie Carroll from autopsies at the Office of the Dade County Medical Examiner, Miami, Florida, and were blindly evaluated for age at death by L.L.K. using the McKern-Stewart [2] system, the Suchey-Brooks 1988 method [4], and the Katz and Suchey 1989 refinement [8] in which the Suchey-Brooks phases are assigned mean ages and standard deviations according to race. This latest refinement of the Suchey-Brooks system was based on the analysis of whites, blacks, and Mexicans autopsied at the Office of the Chief Medical Examiner-Coroner, County of Los Angeles. Since Mexican was not a category of classification for the Dade County sample, only black and white values were used; many of the individuals in the Dade County sample were Hispanic and included with the white sample. Although the Dade County sample represents a somewhat different racial and ethnic mix from the large Los Angeles County sample upon which the Suchey-Brooks system was based, it nevertheless is a sample appropriate to the forensic science situation in another part of the country and may serve as a test of applying that system more broadly. The Dade County sample was taken from 57 black and 59 white males ranging in age from 16 through 89 years. The age distributions of the two groups differed (Fig. 1); the mean age of whites was 50 years, of blacks 41 years. All skeletal samples were obtained consistent with provisions of the Uniform Anatomical Gift Act and state and local regulations.

For both the McKern-Stewart and Gilbert-McKern methods, mean age, standard deviation, and range were based on the total score from all three components. For the Gilbert-McKern system the corrected standard deviations given by Ubelaker [9] were used. For both Suchey-Brooks methods, mean age, standard deviation, and 95% range were based on assigned phase score; however, no 95% range is given for the 1989 refinement for males. Note that for the McKern-Stewart and Gilbert-McKern methods the stated age range for each total score represents the total observed range, but for the Suchey-Brooks methods it is the 95% range. For the test samples no age criteria other

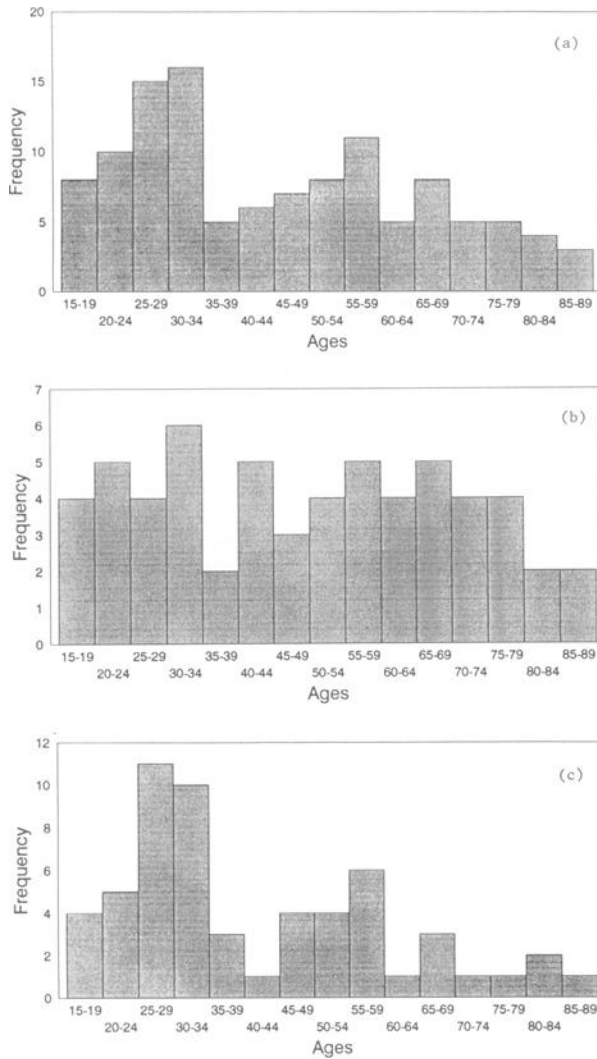


FIG. 1—Age distribution charts: number of individuals in Dade County sample by five-year cohorts for (a) all males, (b) white males, and (c) black males.

than pubic morphology were available for evaluation. Although this circumstance is rarely, if ever, encountered in practice, it does provide a stringent test for the evaluation of aging methods based on the *os pubis*.

For each method, performance was first measured by mean absolute deviations of true age from scored interval means. Two-tailed paired *t*-tests were performed on mean age deviations for McKern-Stewart versus Suchey-Brooks (both combined racial means and where mean age was assigned according to race) for the males and for Gilbert-McKern versus Suchey-Brooks for the females. A second performance test measured the proportion of cases in which the true age fell within ± 1 standard deviation, ± 2 standard deviations, and within the stated age ranges of each scored interval. For the purposes of these tests the Gilbert-McKern top range for the highest possible score (15) was consid-

ered limitless; McKern-Stewart [2] gives limitless top ranges for total scores of 14 and 15. This was not the case for the Suchey-Brooks Phase 6 since theirs was 95% ranges, and one would expect at least 2.5% of cases to lie above them. Although this approach is useful for a statistical evaluation, in practice the forensic anthropologist would consider that the highest possible score would include the most advanced ages.

Results

Tables 1 and 2 present the results for females and males. For females the average absolute deviation of the true age from the mean phase age was 16.1 years for the Suchey-Brooks system, and the average deviation of true age from total score mean age was 14.6 years for the Gilbert-McKern system. The difference between these means is statistically significant by a paired *t*-test ($p < 0.006$) favoring Gilbert-McKern. On the other hand, the proportion of cases in which true age fell within ± 1 standard deviation, ± 2 standard deviations, or within the stated age range was much higher for Suchey-Brooks. This results in large part from the larger standard deviation and range intervals for the Suchey-Brooks system.

For the males the McKern-Stewart deviation of true age from the total score interval mean was 15.8 years. The Suchey-Brooks system for all races pooled together gave an average deviation of 11.1 years; when phase means were assigned according to race, 10.5 years. The differences between the McKern-Stewart mean deviations and both Suchey-Brooks systems were significant ($p < 0.0000$) by a two-tailed paired *t*-test, favoring Suchey-Brooks. The mean deviations between the Suchey-Brooks combined race and

TABLE 1—Mean absolute deviations of true age from interval mean and proportion of test sample within ± 1 , ± 2 standard deviations (SD) and stated range for Los Angeles County females (N = 202) using the Gilbert-McKern (G-M) and Suchey-Brooks (S-B) methods.

Method	Mean Deviation in Years	Percentage of Sample		
		± 1 SD	± 2 SD	Stated Range ^a
G-M	14.6	26	39	46
S-B	16.1	43	70	81

^aFor Gilbert-McKern (G-M) 100% range; for Suchey-Brooks (S-B) 95% range.

TABLE 2—Mean absolute deviations of true age from interval mean and proportion of test sample within ± 1 , ± 2 standard deviations (SD) and stated age range for Dade County males using the McKern-Stewart (M-S) and Suchey-Brooks (S-B) methods.

Method	Race	N	Mean Deviation in Years	Percentage of Sample		
				± 1 SD	± 2 SD	Stated Range ^a
M-S	all	116	15.6	27	44	68
	white	59	17.6	19	42	75
	black	57	13.5	35	46	61
S-B 1988	all	116	11.1	58	75	79
	white	59	12.8	49	69	76
	black	57	9.4	67	81	82
S-B 1989	all	116	10.5	47	83	...
	white	59	11.4	46	83	...
	black	57	9.6	49	82	...

^aFor McKern-Stewart (M-S) 100% range; for Suchey-Brooks (S-B) 95% range. Ranges not published for Suchey-Brooks 1989 refinement by race [8].

racially specific refinement was also significant at the 0.03 level, favoring the Katz and Suchey refinement. The percentage of cases falling within ± 1 and ± 2 standard deviations was consistently much higher for both Suchey-Brooks systems than for McKern-Stewart. For percentage of cases falling within ± 2 standard deviations the racially specific refinement of Suchey-Brooks was superior to the pooled race figures, although this was not so for ± 1 standard deviation. For all methods the performance was better for blacks than for whites by all criteria.

Discussion

By the standards of evaluation used in this study, all the aging methods based on the *os pubis* proved disappointing in regard to both accuracy and precision. Not surprisingly, the deviations tended to be lower for younger individuals and higher for the aged. In none of the methods did the proportion of cases falling within a designated interval (± 1 standard deviation, ± 2 standard deviations, stated age range) even closely approach the proportion expected from the original database samples (68, 95, and 95 to 100%, respectively). However, the Suchey-Brooks methods fared much better in this respect for both males and females.

The relatively poor performance of the McKern-Stewart method probably reflects the restricted nature of the database on which it was built: very predominantly, young, white males. This does not accurately reflect the population from which the test sample was drawn. For most populations the McKern-Stewart standard deviations are unrealistically constricted. For example, in one case the actual age was 20; the estimated age was 19.0, but this did not fall within ± 1 standard deviation. The Gilbert-McKern database, although demographically less restricted than that of McKern-Stewart, numbered only 103. The larger standard deviations of the Suchey-Brooks methods doubtless reflect the larger and more diverse samples on which they were based.

The differences in database samples among the methods may well explain the smaller deviation of actual age from score mean age for Suchey-Brooks versus McKern-Stewart. The relatively better performance of Gilbert-McKern versus Suchey-Brooks on this measure is not so easily explained. Perhaps the individual idiosyncrasies of the observer were significant, but these are not subject to easy analysis. Note that the observer had considerable experience dealing with both of the three-component systems, but had never before used either Suchey-Brooks method. The greater accuracy of estimated age for male than females pubes supports the conventional wisdom that female pubes are far more variable and therefore harder to evaluate. The higher variability in female pubes presumably reflects considerable variability in complex obstetrical factors that influence symphyseal morphology.

A fatigue factor may also play a role when large samples are being evaluated since missed estimates tended to cluster in time. Although not included as part of this research design, repeated observations using the same standards and samples would serve to evaluate intra-observer error. There is another hint that large samples may adversely affect the performance of the observer. A small sample of female pubes ($N = 27$; age range of 18 to 96) from Dade County was also evaluated by the Suchey-Brooks method. In this case, 63% of estimates fell within ± 1 standard deviation, and only 2 (7%) fell outside of all ranges. This suggests that for forensic science cases these methods are more reliable than the large test samples suggest.

All methods performed much better on younger individuals. For example, for females under the age of 30, the average absolute deviation was 7.1 years, and the true age fell within ± 2 standard deviations 24 of 29 times or 83% of the time with Gilbert-McKern. The average deviation was 5.3 years and the true age fell within ± 2 standard deviations 28 of 29 or 97% of the time for Suchey-Brooks. For males under the age of 35 the average

absolute deviation was 3 years, and the true age fell within ± 2 standard deviations 39 of 50 times or 78% of the time for McKern-Stewart. The average deviation was 4.4 years, and true age fell within ± 2 standard deviations 48 of 50 times or 96% of the time for Suchey-Brooks. A note of caution should be added about interpreting the proportion of cases falling within standard-deviation intervals. The confidence level of 95% for values falling within two standard deviations of the mean requires an assumption of normality, which is probably not a reasonable assumption. Moreover, the width of the various standard deviation intervals and ranges varies considerably across methods. As a result, it is not meaningful to compare directly the proportion of cases falling within different intervals. One might expect, for example, that a method yielding wider intervals would result in a higher proportion of cases falling within the interval. Thus, one should consider jointly the interval widths, the proportion of the sample included in the interval, and the level of confidence one would have that a given age would fall within the corresponding interval. For a given level of confidence, the width of the corresponding age interval is narrower for younger pubes, and the average deviations of true age from the interval means are also smaller. The greater accuracy in estimating the age of black than white males probably reflects in large part the younger ages of the blacks (Fig. 1). The lower percentage of cases falling into ± 1 standard deviation by the Suchey-Brooks racial refinement compared with the original pooled race method may have resulted from different spreads of the standard deviations for some phases.

Several factors regarding the male pubes should be noted. In contrast to the female samples, removal of the symphyseal faces sometimes had resulted in either the lowermost tip missing (27 cases) or the lower $\frac{1}{3}$ to $\frac{1}{2}$ missing (21 cases). Surprisingly, the completeness of the specimen generally had little influence on success of age estimation. Twelve male pubes were inaccurately assessed by all methods and criteria. Of these, only one age was overestimated, and in that case the bottom half of the symphysis was missing. One was noted as very unusual at the time of evaluation. One from a 50-year-old was underestimated to the 20's on the basis of morphology. This individual had undergone below-the-knee amputation of the right leg. The amputation had probably been several years long-standing since it was completely healed, the right thigh musculature had atrophied, and the pubes were very osteoporotic. In contrast, a 26-year-old had undergone a bilateral below-the-knee amputation roughly two years previously. The estimated age of about 23 suggests that the amputation had not been sufficiently long-standing to affect seriously the age estimate. A 51-year-old had been estimated as late twenties, and this man was labeled as a disabled veteran. An 87-year-old whose age was badly underestimated was listed as having organic brain syndrome and at autopsy was 167.6 cm and 29.1 kg (5 ft-6 in. and 64 lbs)—indicating a long period of physical inactivity. Another underestimated 74-year-old had a healed fracture of the right pubic ramus and was listed as having massive healed fractures of both legs; at autopsy he was 165.1 cm and 46.6 kg (5 ft-5 in. and 102 lbs). These cases suggest that serious physical trauma or lack of normal physical activity or both may interrupt the normal sequence of pubic morphological age changes. Two additional cases of significant age underestimation occurred in men who were very underweight at autopsy: 185.4 cm and 62.3 kg (6 ft-1 in. and 137 lbs) and 180.3 cm and 50 kg (5 ft-11 in. and 110 lbs); the latter was known to be a chronic alcoholic. For the remaining four cases of significant underestimation, no unusual characteristics had been noted except that two were overweight.

If one compares the average absolute age deviations reported in this study with those reported for males by Meindl and co-workers [7], specifically 5.0 years, using their revised Todd method for estimating age from the pubis, one might conclude that the Suchey-Brooks method is less accurate than the revised Todd system. However, when drawing from two studies with different designs and samples, true comparisons are very difficult

to make. The Kent State sample, chosen from the Todd collection, included only individuals aged between 20 and 69 years. Only two of 58 (3.5%) of that sample were over the age of 59 years, whereas 30 of 116 (26%) of the much larger Dade County sample were over the age of 59. Moreover, the blind age evaluations of the Todd sample were performed by workers involved in devising the revised method, which was not the case in this study. The present study evaluates the Suchey-Brooks method in the hands of an average osteologist. Also, the original Todd system was based on the Todd collection, so the inaccuracy figures reported by the Kent State group do not really test the system on a completely independent sample. In contrast to the Kent State work, no attempt at seriation of pubes were made in the present study since it is generally not applicable in the forensic science situation; instead, each pubis was evaluated independently. Finally, the accuracy of the known ages of the Todd collection specimens is uncertain.

Conclusions

What are the implications of this study for the practice of forensic anthropology?

1. The most important caution would involve the use of pubic age to exclude possible candidates for identification. One should *not* assume that in 95% of cases true age will lie within ± 2 standard deviations of the mean age for the score, or even within the stated age range. The temptation to eschew the Suchey-Brooks methods because of their much larger age ranges should be resisted, for these come much closer to the reality of practice. The much smaller standard deviations and ranges listed for the three-component systems provide only a false sense of security.

2. Not surprisingly, these data indicate that younger-looking pubes can be evaluated with greater accuracy and precision by any of the methods.

3. By all measures used in this study the Suchey-Brooks method outperformed the McKern-Stewart for male pubes. We cannot think of any circumstances under which McKern-Stewart would be the method of choice. The choice is not so clearcut for the two methods applicable to females. In the forensic science situation the chance of inappropriately eliminating an individual from a list of potential candidates for identification is much lower for Suchey-Brooks, and this would be the most important criterion.

4. If racial affiliation of males is known or strongly suspected, the 1989 Katz and Suchey refinement should be used, especially if the individual is known or thought to be black or white.

5. Reported age estimates based on pubic symphysis morphology should include ± 2 standard deviations.

6. Any evidence of former trauma or debilitation should be taken into account, and the anthropologist may wish to consider whether the age estimate based on pubic morphology should be inflated.

7. Finally, we would emphasize that in actual cases other skeletal age indicators would normally be used in conjunction with pubic aging methods. Multifactorial-based age estimates are important in improving one's chances of arriving at an age range that would include the true age and yet not span four or five decades, even for older individuals. All available age criteria should be considered in actual case analysis.

Acknowledgments

We wish to thank Judy Suchey and Paul Sledzik whose efforts and cooperation contributed substantially to this study.

References

- [1] Brooks, S. and Suchey, J. M., "Skeletal Age Determination Based on the Os Pubis: a Comparison of the Acsádi-Nemeskéri and Suchey-Brooks Methods," *Human Evolution*, Vol. 5, No. 3, 1990, pp. 227-238.
- [2] McKern, T. W. and Stewart, T. D., "Skeletal Age Changes in Young American Males," Technical Report EP-45, U.S. Army Quartermaster Research and Development Center, Environmental Protection Research Division, Natick, MA, 1957.
- [3] Gilbert, B. M. and McKern, T. W., "A Method for Aging the Female Os Pubis," *American Journal of Physical Anthropology*, Vol. 38, No. 1, 1973, pp. 31-38.
- [4] Suchey, J. M., Brooks, S. T., and Katz, D., "Instructions for Use of the Suchey-Brooks System for Age Determination for the Female Os Pubis," instructional materials accompanying female pubic symphyseal models of the Suchey-Brooks system, distributed by France Casting (Diane France, 20102 Buckhorn Rd., Bellvue, CO 80512), 1988.
- [5] Suchey, J. M., "Problems in the Aging of Females Using the Os Pubis," *American Journal of Physical Anthropology*, Vol. 51, No. 3, Sept. 1979, pp. 467-470.
- [6] Suchey, J. M., Wisely, D. V., and Noguchi, T. T., "Aging the Male Os Pubis," presented at the 32nd Annual Meeting of the American Academy of Forensic Sciences, New Orleans, Feb. 1980.
- [7] Meindl, R. S., Lovejoy, C. O., Mensforth, R. P., and Walker, R. A., "A Revised Method of Age Determination Using the Os Pubis with a Review and Tests of Accuracy of Other Current Methods of Pubic Symphyseal Aging," *American Journal of Physical Anthropology*, Vol. 68, No. 1, Sept. 1985, pp. 29-45.
- [8] Katz, D. and Suchey, J. M., "Race Differences in Pubic Symphyseal Aging Patterns in the Male," *American Journal of Physical Anthropology*, Vol. 80, No. 2, Oct. 1989, pp. 167-172.
- [9] Ubelaker, D. H., *Human Skeletal Remains*, Aldine, Chicago, 1978.

Address requests for reprints and additional information to
Linda Klepinger, Ph.D.
Department of Anthropology
University of Illinois
109 Davenport Hall
607 S. Mathews
Urbana, IL 61801