Max M. Houck,[1] M.A., Douglas Ubelaker,[2] Ph.D., Douglas Owsley,[2] Ph.D., Emily Craig,[3] Ph.D., William Grant,[4] M.A., Robert Fram,[1] M.A., Theresa Woltanski,[5] M.A., and Kari Sandness,[2] M.A.

The Role of Forensic Anthropology in the Recovery and Analysis of Branch Davidian Compound Victims: Assessing the Accuracy of Age Estimations*

REFERENCE: Houck MM, Ubelaker D, Owsley D, Craig E, Grant W, Fram R, Woltanski T, Sandness K. The role of forensic anthropology in the recovery and analysis of Branch Davidian compound victims: assessing the accuracy of age estimations. J Forensic Sci. 1996;41(5):796–801.

ABSTRACT: Age-at-death estimations of 44 individuals (27 adults, 17 children) from the Branch Davidian sample were compared with their actual ages. Estimations were evaluated for bias and accuracy for the actual age at death. Although the overall average estimates correlated well with the actual ages ($r = 0.946$), several individuals displayed high residual requiring further analysis and review. These individuals displayed age-related features that did not correspond with the expected morphology for individuals of their ages. Several age estimation techniques scored these individuals with all bias in the same direction. These examples should serve as cautionary reminders that biology does not always correlate with expected outcomes, particularly in such multifaceted traits such as age.

KEYWORDS: forensic science, forensic anthropology, physical anthropology, age-at-death estimation, Branch Davidian, Compound

The anthropological contribution to the investigation of the Branch Davidian incident near Waco, Texas provided recovery and analysis of the human remains at the site (1,2). Part of this analysis was the development of biological profiles, estimations of sex, age, ancestry, stature, and trauma that assisted in the association and identification of the deceased. At the time of the completion of the autopsies, 88 individuals had been examined, 83 of them anthropologically. Forty-one individuals were positively identified by dentition, fingerprints, or premortem X-rays during the autopsy process (1,2). Biological profiles had been developed for the remaining 42 individuals who were later identified by polymerase chain reaction (PCR)-based DNA typing (3); two additional individuals from those initially identified are included in this sample as age estimations were made before identification. This provided a total sample of 44 subsequently identified individuals. This confirmation of identity provided an opportunity to check the accuracy and bias of various age estimation methods against a "modern" sample of known age. In typical casework, forensic anthropologists deal with only one or at most several individuals, and these stochastic observations do not allow for evaluation of methodologies based on large samples. Ideally, one would like to have a recent population with which to compare contemporaneous individual observations (4). The sample of DNA-identified individuals provided a recent population against which methods of age estimation could be assessed. Each individual's age was verified through public documents such as birth records or driver's licenses. Of the 27 adults, 18 were female and nine were male; as a result of the distribution of the ages (Fig. 1) and the discontinuity between 13 and 17 years, 17 years and greater was used as the demarcator of adulthood. Eleven of the children were female and seven were

Frequency	"Tens"	"Ones"
12	0	111122222444
16	0	5789
17	1	3
19	1	78
20	2	3
(7)	2	557[7]788
17	3	0011133
10	3	567
7	4	023
4	4	78
2	5	
2	5	5
1	6	1

FIG. 1—*Stem-and-leaf of all ages of Branch Davidian sample. Each value was separated into its first ("tens") and second ("ones") digits with the first digits broken into groups of 0–4 and 5–9 and the second digits listed individually. For example, the third stem reads "13" and the fourth stem reads "17, 18." Frequency is shown in the first column, ascending to the median group (the median value is in brackets), then descending.*

Received for publication 11 sept. 1995; revised manuscript received 2 Feb. 1996; accepted for publication 5 Feb. 1996.

[1]Forensic examiner and supervisory special agent, respectively, FBI Laboratory, Hairs and Fibers Unit, Washington, DC.

[2]Curators and research assistant, respectively, National Museum of Natural History, Department of Anthropology, Smithsonian Institution, Washington, DC.

[3]Forensic anthropologist, Medical Examiner's Office, Frankfort, KY.

[4]Anthropologist, US Army Central Identification Laboratory, Hickam Air Force Base, Honolulu, HI.

[5]Criminalist 2, Tarrant County Medical Examiner's Office Criminalistics Laboratory, Fort Worth, TX; present address: US Army Central Identification Laboratory, Hickman AFB, Honolulu, HI.

*This is publication 95-16 of the Laboratory Division of the Federal Bureau of Investigation. Namees of commercial manufacturers are provided for identification only, and their inclusion does not imply endorsement by the Federal Bureau of Investigation.

male. The distribution is bimodal and shows the majority of children to be fewer than five years of age and the majority of adults to be in their late-20s and early-30s with the median age being 27 years.

Methods

No single method of estimation of age at death is likely to reflect accurately the changes that accumulate in the skeleton with chronological age, therefore, no single method should be used for skeletal populations including forensic casework (4,5). Methods of age estimation were chosen based upon their published accuracy, the skeletal elements from each individual that were sufficiently preserved for analysis, the familiarity each anthropologist had with the particular technique, and the appropriateness of the technique to the initial estimated age range of the individual. This approach proved to be the most effective way for multiple anthropologists to examine a large number of remains representing a variety of taphonomic conditions. Because of the variability in the preservation of the remains, not all methods of age estimation were applicable to all individuals; this precluded detailed intercomparison of the remains, age estimation methods, and results.

The collective experience of the anthropologists, additionally, played a large role in the interpretation of the remains. Most of the remains were evaluated as a "group effort," although some individuals were assessed by one person with a subsequent check by the other investigators. The following techniques were used: long bone length (4,6,7), epiphyseal appearance and union (7–9), dental development (10,11), pubic symphyseal morphology (12–15), auricular surface morphology (16), sternal rib morphology (17–19), and degenerative changes (9,20–23). The condition of each individual was recorded on the University of Tennessee Anthropology Database inventory sheets. On this form, skeletal elements were scored as present, fragmentary, or absent postmortem. Notes concerning the general condition of the remains as well as age estimations were entered directly into a laptop computer with word-processing software. This facilitated daily reviews of work, reduced typing for later report generation, and allowed for searches by keywords or individuals.

A probable age range was developed for each individual based on our interpretation of the features presented at autopsy. These initial estimates were later refined, either in the morgue or after cleaning of an age-related specimen. Simple averages of estimated ranges of age at death were calculated for each individual because very little improvement is gained by applying complex statistical procedures to this type of analysis (5).

An estimator's accuracy refers to the distance a particular estimated value lies, on average, from the true value of the parameter being measured, in this case, age at death. The accuracy of an estimator can be evaluated on the basis of its mean square error (MSE). The smaller the MSE of an estimate, the greater its accuracy (24). Using statistics such as mean, standard deviation, and variance can result in a misleading view of data because they rely on squared deviations from the mean, cases farther from the mean not only increase the sum of squares but also do so at an increasing rate. This makes these statistics nonresistant to being disproportionately influenced by a small proportion of the total number of observations (25). Therefore, as often as possible resistant statistics, such as the median and techniques of exploratory data analysis (25,26), were used to address the question of accuracy and bias. In simple problems involving a single set of measurements of one variable, we can think of each measurement as having two components, the indicator of central tendency and the residual.

Practically speaking, the residual is the deviation from a mean, median or other location parameter (27). A simple method of identifying outlying results is to address any measurements greater than 1.5 times the midspread that is the distance between the third quartile and the first quartile (25,27). Also, the distance between the observed statistic and the estimated value in units of the variability of the statistic, that is, correlation and regression, was measured.

Results

The sample of estimated ages at death showed close statistical similarities with the actual sample ages (Table 1) and the estimated ages also patterned closely with the known samples (Fig. 2). Calculation of correlation, analysis of variance, and regression demonstrated a good correlation between the known and estimated samples ($r = 0.946$) and accounts for almost 90% of the variation in the estimated ages with no evidence for a lack of fit ($P > 0.1$, $P_{sample} = 0.1274$) (Fig. 3). This close correlation can be seen in a scatterplot of the data with a Tukey line imposed upon it (Fig. 4). A Tukey line is a resistant analog of a least-squares regression line and is constructed as follows: a line is drawn from the median values of the first third of the x and y data to the median values of the last third of the x and y data values; this line is then moved parallel to its original position until half the observations are above the line and half are below it (25). Tukey lines serve as easy-to-generate, outlier-resistant analogs of fitted least-square regression lines. As a result of the regression equation, several estimated ages showed large standard residuals that required explanation (Table 2). These outliers will be addressed later.

Bias and inaccuracy were calculated for each individual and then summarized for the following demographic groups: all individuals, male adults, female adults, male children, female children, all adults, all children, all males, and all females (Tables 3 and 4). Bias and inaccuracy were also calculated by decade for males and females (Table 5). The overall bias was plotted (Fig. 5); probable outliers fall between 1.5 and 3 times the midspread are denoted with an "0." Note that the greatest bias values in adults correspond with the five unusual observations found earlier.

Discussion

Overall, the estimation of age at death for the Branch Davidian sample was accurate showing a high correlation between the known and estimated samples and low bias for all groups. This indicates that, although no one age estimation method proved itself most valuable, a consensus opinion derived from a variety of techniques can be very accurate. The most erroneous statistics calculated for this sample are that of the bias and inaccuracy based on the median

TABLE 1—*Descriptive statistics for ages at death of Branch Davidian sample. S.D. = standard deviation, Q1 = lower quartile, Q3 = upper quartile.*

Age	Mean	Median	S.D.	Minimum	Maximum	Q1	Q3
Actual age	22.18	26	16.83	1	61	4	33
Average estimated age	22.24	25	16.27	0.9	55	4.87	34.87

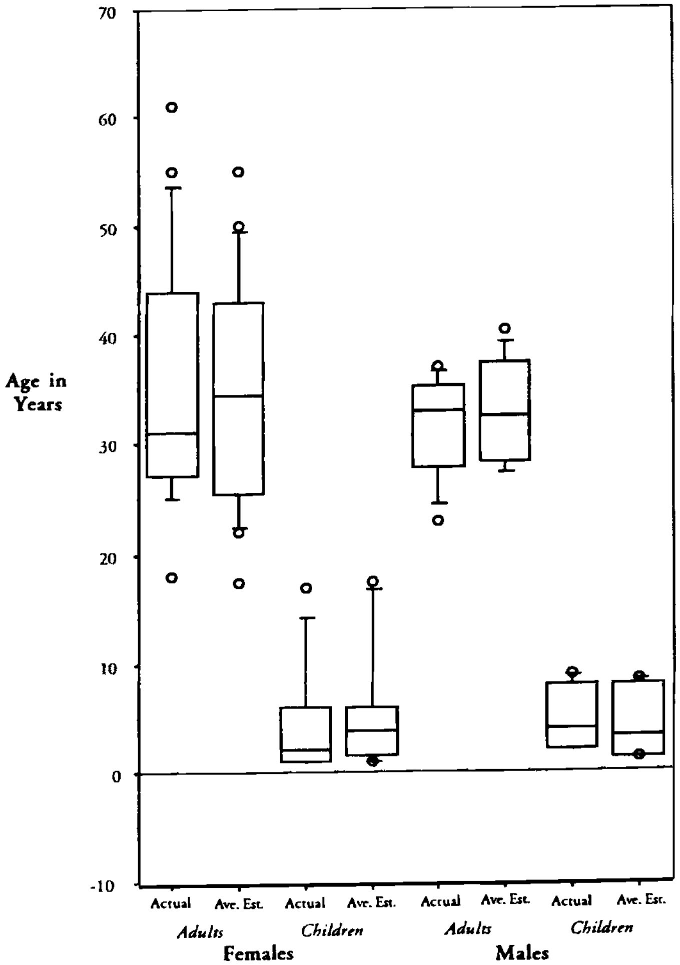

FIG. 2—*Boxplot of actual and estimated ages at death of Branch Davidian sample.*

Correlation

Predictor	Coefficient	Standard Deviation.	t-ratio	P
Constant	0.414	1.418	0.29	0.772
Ave. Est	0.979	0.052	18.94	0.000

Estimate of σ = 5.512
Square of correlation coefficient (R^2) = 89.5%
R2 adjusted for degrees of freedom = 89.3%

Analysis of Variance

	Degrees of Freedom	Sum of Squares	Mean Square
Regression	1	10904	10904
Error	42	1276	30
Total	43	12181	

FIG. 3—*Results of correlation and regression calculation on actual and estimated ages at death for Branch Davidian sample.*

values for female adults, 3.5 and 8.0 years, respectively, which are small margins of difference when compared with other published evaluations of documented samples (for example, see Ref 5). Adult males also had high bias and inaccuracy but were scored more accurately than females in all cases; this general relationship has

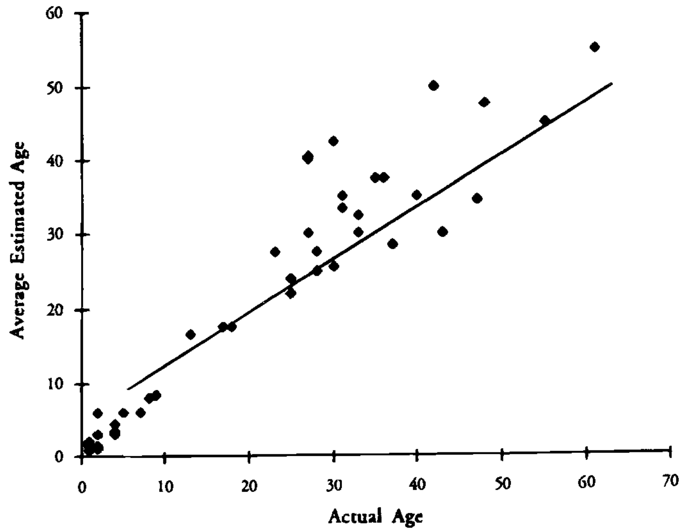

FIG. 4—*Scatterplot of average estimated age versus actual age of Branch Davidian sample with Tukey "regression" line.*

been noted elsewhere (15). The children's estimates were the most accurate and this is probably because dentition was the primary mode of assessing age at death. As published, the method of Moorrees, Fanning, and Hunt (10,11), at its worst, provides accuracy to ±1 year. Estimations on female children fared worse than those on similarly aged male children. Given the general lack of research into age and prepubertal sex differences for children, no reasonable explanations can be provided at this time.

The difference seen between the mean and median inaccuracies for all individuals may be due to statistical effects from outlying values. The value of the median inaccuracy for all individuals may be a more realistic value than the mean because a wide range of ages (1 to 61 years) distributed bimodally is being investigated.

Several outliers defined by large residuals (bias) indicate that age estimations on an individual-by-individual basis can be occasionally quite inaccurate. Table 2 lists the individuals with the largest deviations from the expected age at death. In an effort to explain why these individuals in particular were evaluated less accurately than other individuals of the same age bracket, the original notes taken at autopsy were reviewed. All five individuals were heavily charred and fragmented. Samples for postautopsy cleaning and age estimation were taken from four of the five individuals. In all five cases, initial age estimations made at autopsy were not far from the final age estimation made after cleaning: all final age estimations fell within the initial estimated ranges. Interestingly, the notes from one of the autopsies read that "this individual demonstrates both youthful and older characteristics" (authors' personal notes).

Another factor in the deviations could be that, for some of the age estimation methods, a much narrower range than the 95% confidence interval range was reported. In the interest of providing the optimum age range (i.e., the narrowest range with the maximum confidence) to help identify individuals, the 95% confidence interval was often too wide to be diagnostic.

If we hold one indicator of age as a constant (the public symphysis) and compare the descriptions across several variations of a single technique (Todd, McKern, and Stewart, Modified Todd) we still find that all tested methods tend to be misleading for these individuals. For example, the description of the pubic symphyseal surface of Observation 41, a 47-year-old female, reads, "very slight

TABLE 2—*Unusual observations of large standard residuals from the age at death estimations for Branch Davidian sample.*

Observation	Average Estimated Age	Actual Age	Fit	Fit S.D.	Residual	Standard Residual
23	40.0	27	39.557	1.238	−12.557	−2.34
25	40.5	27	40.046	1.257	−13.046	−2.24
28	42.5	30	42.003	1.336	−12.003	−2.24
40	30	43	29.771	0.923	13.229	2.43
41	34.5	47	34.175	1.045	12.825	2.37

TABLE 3—*Bias of estimated age at death for various demographic groups of the Branch Davidian sample.*

Statistic	Males	Females	All
Mean:			
Adults	1.33	−0.94	−0.15
Children	−0.6	0.9	0.38
All	0.56	−0.19	0.06
Median:			
Adults	0.5	3.5	−0.15
Children	−0.75	1.75	0.25
All	0.5	−1.0	−0.75

TABLE 4—*Inaccuracy of estimated age at death for various demographic groups of the Branch Davidian sample.*

Statistic	Males	Females	All
Mean:			
Adults	4.17	6.0	5.37
Children	0.60	1.14	0.96
All	2.74	3.69	6.00
Median:			
Adults	2.5	8.0	3.5
Children	0.63	0.75	0.63
All	1.0	2.75	2.5

TABLE 5—*Bias by decade of estimated age at death of the Branch Davidian sample.*

		Bias	
Decade	*N*	Males	Females
0–10	16	−0.1	0.36
10–20	3		−1.17
21–30	8	2.83	2.86
31–40	10	−2.33	2.17
41–50	5		2.00
51–60	1		10.00
61–70	1		6.00

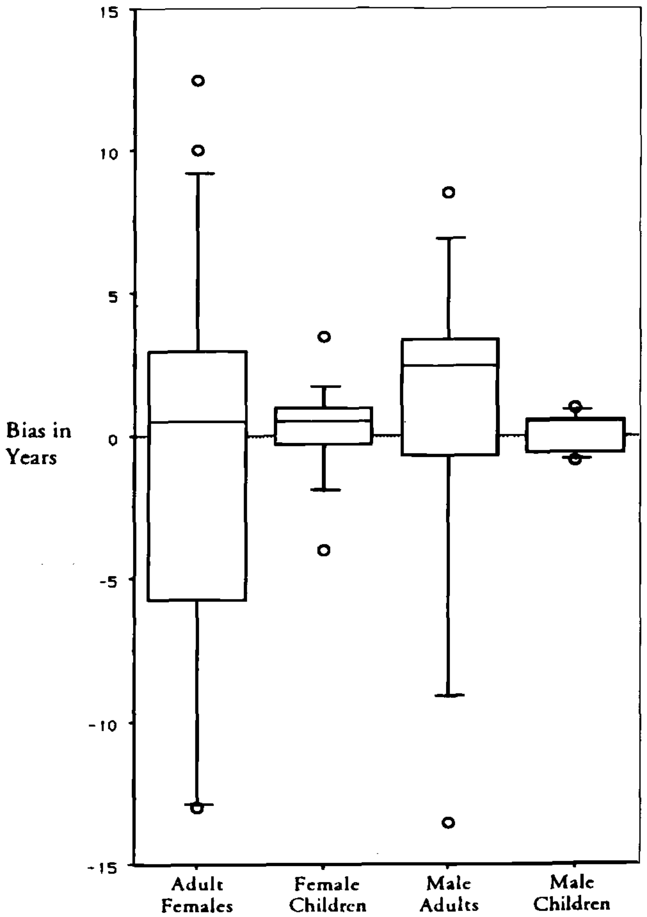

FIG. 5—*Bias of age at death estimations of the Branch Davidian sample.*

vestigial billowing. The ventral rampart is nearly complete with slight rim formation on the medial surface" (authors' personal notes). This observation yields the following age ranges: Todd, Phase VI, 30 to 35 years, mean = 32.5 years; Suchey-Brooks, Phase 4, 27.3 to 49.1 years, mean = 38.2 years; Modified Todd, Phase C, 22 to 50 years, mean = 36 years. Our "composite" method yielded an age range of 30 to 39 years, mean of 34.5 years. These data and the data for other individuals are presented in Table 6. If we compare these individuals with others in the same sex and age cadres, they still stand out as exceptions (Table 7).

In a paper evaluating pubic symphysis aging methods, Suchey, Wiseley, and Katz (28) note that the original Todd method tends to overage a "modern" multiracial sample from Los Angeles. They note that "older patterns (IX and X) can occur in males in their twenties" (28:45). The IX phase corresponds to a 45 to 50 age range and the X range to 50 or more years. This is approximately what may have occurred with Observations 25 and 28, both males with known ages toward the end of the second decade who were scored at being over 40 years. Suchey, Wiseley, and Katz further note that age 40 and above is a good cutoff to expect a reasonable degree of accuracy. This, coupled with an intrinsic general lack of accuracy of age estimation methods for females (15), could account for the errors that occurred with Observations 23, 40, and 41. It is interesting to note that Observations 40 and 41 were both under-aged to roughly the same degree. It is reasonable to assume these individuals clearly presented characteristics that were not consistent with the predicted traits for persons of their sex and chronological age. Any errors were amplified by inherent bias and inaccuracy in the age estimation methods used. By using multiple

TABLE 6—*Mean estimated ages at death using various pubic symphysis techniques on individuals with high residuals.*

Observation Number	Sex	Todd	McKern-Stewart	Suchey-Brooks (M)	Suchey-Brooks (F)	Modified Todd	Actual Age
23	F	42.5	n/a	n/a	38.2	45+	27
25	M	42.5	41	36.8	n/a	45+	27
28	M	42.5	41	36.8	n/a	45+	30
40	F	28.5	n/a	n/a	30.7	24.5	43
41	F	32.5	n/a	n/a	38.2	36	47

TABLE 7—*Outliers of estimated ages at death compared with sex and age cadres. Boxed values are outliers discussed in text.*

Observation Number	Actual Age	Average Estimated Age	Bias
Males:			
	25	24	1
	25	22	−3
23	27	40	−13
	27	30	3
	28	25	−3
	30	25.5	−4.5
...	...	...	...
	42	50	8
40	43	30	−13
41	47	34.5	−12.5
	48	47.5	−0.5
	23	27.5	4.5
Females:			
25	27	40.5	13.5
	28	27.5	−0.5
28	30	42.5	12.5
	31	33.5	2.5
	33	32.5	−0.5
	33	30	−5

methods of age estimation based on various anatomical features, we hope that this type of error could be reduced to an acceptable level.

This evaluation of age estimation accuracy yields a variety of common sense and the important cautionary remarks concerning age estimations. Multiple methods of age estimation based on a variety of age-related structures should be used on an individual. This should reduce the errors inherent in each method and any misleading features presented by any one individual. Within certain boundaries, different areas of the skeleton can age at different rates. Although it is natural to assume that an age estimation using, for example, the auricular surface should correlate highly with that derived from the sternal rib end, this may not necessarily be so. For this sample, experienced evaluation with multiple methods was the best way of producing an overall age assessment. Whenever possible, the age-related skeletal changes of an individual should be compared with a modern population. Experience is a key aspect of this type of evaluation, and not simply experience with the method at hand, but also with skeletal biology and anatomy. The accuracy of the age estimations for this sample depended in large part on the experience of the anthropologists involved. Having multiple observers often proved useful in producing age estimates particularly with difficult samples or equivocal indicators.

Although science often hopes to develop "pure" techniques that can be accurately used by novice and master alike, experience is nonetheless the determining feature of any evaluative process. Technique alone cannot carry the day. An "insistence upon exactness has to bow to the requirements of adequacy" (29), and those requirements can be decided upon only after sufficient experience has been accumulated. Because one of the inviolable laws of biology is that individuals and populations vary (30), this variation must be taken into account with every case examined. The characterization of this variety and rendering it intelligible is the goal of all biological sciences, particularly forensic anthropology. By analyzing modern samples and reflecting on the results, we can continue to refine and define the appropriate application of our techniques.

Acknowledgments

The authors would like to thank Dr. Mary W. Tungol and Dr. Joe DiZinno for their comments on early drafts of this paper.

References

(1) Ubelaker DH, Owsley DW, Houck MM, Craig E, Grant W, Woltanski T, et al. The role of forensic anthropology in the recovery and analysis of Branch Davidian compound victims: recovery procedures and characteristics of the victims. J Forensic Sci 1995; 40(3):335–40.
(2) Owsley DW, Ubelaker DH, Houck MM, Sandness KL, Grant WE, Craig EA, et al. The role of forensic anthropology in the recovery and analysis of Branch Davidian Compound victims: techniques of analysis. J Forensic Sci. 1995;40(3):341–48.
(3) DiZinno J, Fisher D, Barritt S, Clayton T, Gill P, Holland M, et al. The Waco, Texas incident: the use of DNA analysis to identify human remains. Scottsdale, AZ: IR: Proceedings from the 5th International Symposium of Human Identification, 1994.
(4) Ubelaker DH. Human skeletal remains. Washington, DC: Taraxacum, 1989.
(5) Saunders SR, Fitzgerald C, Rogers T, Dudar C, McKillop H. A test of several methods of skeletal age estimation using a documented archaeological sample. Can Soc Forensic Sci J. 1992;25(2):97–117.
(6) Fasekas IG, Kosa F. Forensic fetal osteology. Budapest: Akademiai Kaido, 1978.
(7) Stewart TD. Essentials of forensic anthropology, especially as developed in the United States. Springfield, IL: CC Thomas, 1979.
(8) Krogman WM, İşchan MY. The human skeleton in forensic medicine. 2nd ed. Springfield, IL: CC Thomas, 1986.
(9) McKern TW, Stewart TD. Skeletal age changes in young American males. Technical Report EP-45. Natick, MA; Headquarters, Quartermaster Research and Development Command, 1957.
(10) Moorrees CFA, Fanning EA, Hunt Jr. EE. Formation and resorption of three deciduous teeth in children. Am J Phys Anthropol 1963;21:205–13.
(11) Moorrees CFA, Fanning EA, Hunt Jr. EE. Age variation of formation stages for ten permanent teeth. J Dent Res 1963;42(6): 1490–502.

(12) Todd TW. Age changes in the pubic bone: I, the male white pubis. Am J Phys Anthropol 1920;3(3):285–334.
(13) Todd TW. Age changes in the pubic bone. Am J Phys Anthropol 1921;4(1):1–70.
(14) Katz D, Suchey JM. Age determination of the male os pubis. Am J Phys Anthropol 1986;69:427–35.
(15) Suchey JM. Problems in aging females using the os pubis. Am J Phys Anthropol 1979;51:467–70.
(16) Lovejoy C, Meindl RS, Pryzbeck TR, Mensforth R. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. Am J Phys Anthropol 1985;68:15–28.
(17) İşchan MY, Loth SR, Wright RK. Age estimation from the rib by phase analysis: white males. J Forensic Sci 1984;2(6):1094–1104.
(18) İşchan MY, Loth SR, Wright RK. Age estimation from the rib by phase analysis: white females. J Forensic Sci 1985;30:853–63.
(19) İşchan MY, Loth SR, Wright RK. Racial variation in the sternal extremity of the rib and its effect on age determination. J Forensic Sci 1987;32:452–66.
(20) Stewart TD. The rate of development of vertebral osteoarthritis in American whites and its significance in skeletal age identification. The Leech 1958;28(3–5):144–51.
(21) Hansen G. Die Altersbestimmung am proximalen Humerus-und Femurende in Rahmen der Identifizierung menschlicher Skelettreste. Wiss Z Humboldt Univ Berl Math Nat Wiss Reihe 1953;1(3):1–73.
(22) Schranz D. Age determinations from the internal structure of the humerus. Am J Phys Anthropol 1959;17(4):273–8.
(23) Hrdlicka A. Practical anthropometry. 2nd ed. Philadelphia: The Wistar Institute of Anatomy and Biology, 1939.
(24) Levy PS, Lemeshow S. Sampling of populations. New York: John Wiley and Sons, 1991.
(25) Hartwig F, Dearing BE. Exploratory data analysis. Newbury Park, CA: Sage Publications, 1979.
(26) Tukey JW. Exploratory data analysis. Reading, MA: Addison-Wesley, 1977.
(27) Mosteller F, Fienberg SE, Rourke REK. Beginning statistics with data analysis. Reading, MA: Addison-Wesley, 1983.
(28) Suchey JM, Wiseley DV, Katz D. Evaluation of the Todd and McKern-Stewart methods for aging the male os pubis. In: Reichs KJ, editor. Forensic osteology. Springfield, IL: CC Thomas Pr, 1986;33–67.
(29) Barrett W. The illusion of technique. Garden City, NY: Anchor/Doubleday Press, 1978.
(30) Mayr E. The growth of biological thought. Cambridge, MA: The Belknap Press of Harvard University Press, 1982.

Address requests for reprints or additional information to
Max M. Houck
FBI Laboratory
Hairs and Fibers Unit, Rm. 3931
10th and Pennsylvania Ave., NW
Washington, DC 20535