Kathy L. Gruspier,[1] *B.A., M.A. and Grant J. Mullen,*[2] *M.A.*

Maxillary Suture Obliteration: A Test of the Mann Method

REFERENCE: Gruspier, K. L. and Mullen, G. J., **"Maxillary Suture Obliteration: A Test of the Mann Method,"** *Journal of Forensic Sciences,* JFSCA, Vol. 36, No. 2, March 1991, pp. 512–519.

ABSTRACT: The degree of obliteration of the maxillary sutures has recently been presented as a useful technique for predicting the age at death of adult skeletal remains. A replicative study of the method on a sample of male skeletons of verified age was undertaken to test its accuracy. Results indicated that the variability in predicted age, as reflected by the total suture score, was too great to recommend use of the method for age determination in individual forensic science cases.

KEYWORDS: physical anthropology, human identification, musculoskeletal system, age at death, interobserver error, linear inverse predictive model, maxillary sutures, obliteration

Recently, three important papers have dealt with the technique of predicting age from the obliteration of the maxillary sutures in humans [*1–3*]. Mann et al. [*1*] assert that the evidence of maxillary suture obliteration may be used to define broad age categories and corroborate the estimates of other aging techniques. Mann et al. [*1*] also suggest that the method could be used to sort commingled remains. Mann's larger study [*2*] sought to explore the use of this information further by generating linear inverse predictive models based on the sum of individual suture scores. In the larger study [*2*], Mann restates these uses of the refined method and, further, cautiously suggests its use when only the maxilla is present. Finally, the results of these studies were presented to the forensic science community by Mann and Jantz [*3*] as a useful technique for the aging of skeletal remains based on the obliteration of the maxillary sutures.

The elegance and seeming simplicity of the method raises two concerns in the mind of the forensic anthropologist: is the method replicable, and what are the quantifiable limits of its accuracy? These concerns were deemed sufficient reason to attempt a replication of the method using the J. C. B. Grant Collection, currently housed at the University of Toronto, Ontario, Canada.

Purpose

The purpose of this investigation was to accomplish the following:

(*a*) to assess interobserver error in the collection of the data particular to this method and

Received for publication 13 March 1990; revised manuscript received 10 May 1990; accepted for publication 15 May 1990.

[1]Graduate student and lecturer, Department of Anthropology, University of Toronto, Toronto, ON, Canada.

[2]Graduate student, Department of Anthropology, Erindale College, University of Toronto, Mississauga, ON, Canada.

(*b*) to determine the relative accuracy of the method of Mann [*2*] when applied to a different population.

Materials and Methods

The J. C. B. Grant Collection is a large sample (n = 200) of anatomy room specimens. The skeletons were collected by Dr. J. C. B. Grant of the Department of Anatomy, University of Toronto, between the 1930s and 1950s. It comprises a majority of males (n = 175), most of which are older than the fourth decade and all but one of which are white. Most of the individuals appear to have been transients, migrant laborers, and recent immigrants, typical of anatomy room populations of the time. The information supplied with each skeleton is the following: the name (if known), date and place of death, cause of death, and recorded age at death. Dr. Grant seems to have been concerned about the integrity of the collection, since evidence was recorded which questioned the correctness of the cause of death and the age at death. Verification of the age at death was through investigation of the background and hospital records of each individual. When only an oral history was available, the lucidity of the patient was recorded.

Within the entire collection, 123 maxillae were suitably undamaged for inclusion in the first interobserver error test sample, which consisted of 83 males and 9 females, for a total of 92 individuals of verified age. The remaining 31 maxillae were later found to have no documented age or a non-verified age, and were then excluded from the analysis. The females were excluded from the second interobserver error test because of an inadequate sample size. The final sample comprised 83 males.

The methods followed in this investigation were those presented by Mann [*2*]. The maxillary sutures examined were the incisive, anterior median palatine, posterior median palatine, and transverse palatine. Each suture was measured with a standard dial caliper from the points defined by Mann [*2*]. The proportion of obliteration was determined by dividing the measured amount of obliteration by the total length of the suture. Obliteration was defined as "any segment of a suture no longer visible, or visible but flush with the adjoining surfaces" [*2*]. The proportion of obliteration for each of the four sutures was summed to give a total suture score. This value was used in Mann's [*2*] inverse prediction formula to obtain a predicted age at death.

All but two of the Grant collection skulls were sectioned slightly lateral to the sagittal plane, necessitating measurement of the incisive and transverse sutures on the side not sectioned. This was contrary to the instructions presented by Mann, who stated that the portion of suture exhibiting the most obliteration should be measured [*2*]. However, since there could be no control for the effects of this modification, it was not included as an analysis variable. The anterior and posterior median palatine sutures were not affected by the sagittal sectioning.

The authors did not record the age and sex of the individuals until data collection was completed.

Analysis Procedure and Results

Analysis of Interobserver Error

Two procedures were established for the evaluation of interobserver error:

1. Spearman rank correlation coefficients were calculated for paired scores obtained by each observer. The 95% confidence limits were calculated to ascertain the range of accuracy being assessed.

2. The Wilcoxon Paired Samples Test was used to evaluate the pairwise equality of the results obtained by the two observers. The procedures presented by Zar [*4*] were utilized for the normal approximation when dealing with zero differences.

Tables 1*a* and 1*b* and 2*a* and 2*b* present the results of two replications of interobserver error testing.

Application of the Model

The inverse prediction model of Mann [2] was applied to the resultant two correlated databases (Observers 1 and 2). Both the graphical presentations and the calculated percentage accuracies were compared with the data of Mann [2].

The percentages of accuracy of the predicted age within certain ranges of the known age were calculated and are presented in Table 3.

Two graphical presentations were best able to show the relationship between the known age and the predicted age. Graphs depicting age versus predicted age and age versus deviation (Figs. 1 through 6) are presented for three sets of data—two data sets generated from the Grant collection (Observers 1 and 2) and the third from Mann's data.

Discussion

The first purpose in this investigation was to study the effects of interobserver error. At a theoretical level, there are two facets to interobserver error in this particular study. The first of these describes error arising from two investigators utilizing the same method on the same population. This error is knowable and quantifiable. The second facet is not quantifiable in that it arises as a result of interpretation of the method as described by the original investigator.

In an attempt to quantify the first facet of error, the Spearman rank correlation coefficient and the Wilcoxon paired samples statistics were performed. The results of the first replication indicated a high interobserver error by low correlation of the data recorded by each observer (Tables 1*a* and 1*b*). At this point, four possible sources of

TABLE 1*a*—*Interobserver Error Test 1 results (n = 123), Spearman correlation coefficients and confidence intervals.*

Suture	r_s	95% Confidence Interval
I	0.633	$0.5095 \leq r_s \leq 0.7311$
A	0.557	$0.4170 \leq r_s \leq 0.6712$
P	0.573	$0.4366 \leq r_s \leq 0.6837$
T	0.687	$0.577 \leq r_s \leq 0.7739$
Total	0.635	$0.5124 \leq r_s \leq 0.7325$

TABLE 1*b*—*Interobserver Error Test 1 results (n = 123), Wilcoxon paired samples test results.*

Suture	$\Sigma T+$	$\Sigma T-$	Z^a
I	1285	670	1.29
A	6299	949	6.79
P	4592.5	1493.5	4.10
T	5725.5	612.5	6.19
Total	6939.5	533.5	7.98

[a]The first test is not significant at $P = 0.05$; the latter four tests are significant at $P = 0.01$.

TABLE 2*a*—*Interobserver Error Test 2 results (n = 83), Spearman correlation coefficients and confidence intervals.*

Suture	r_s	95% Confidence Interval
I	0.829	$0.7658 \leq r_s \leq 0.8787$
A	0.737	$0.6387 \leq r_s \leq 0.8110$
P	0.557	$0.4170 \leq r_s \leq 0.6712$
T	0.739	$0.6434 \leq r_s \leq 0.8110$
Total	0.787	$0.7059 \leq r_s \leq 0.8483$

TABLE 2*b*—*Interobserver Error Test 2 results (n = 83), Wilcoxon paired sample test results.*

Suture	$\Sigma T+$	$\Sigma T-$	Z^a
I	237	323	0.40
A	1580	1311	0.72
P	1325	836	1.27
T	896	1142	1.31
Total	1848	1385	1.06

[a]The five tests are not significant at $P = 0.05$.

TABLE 3—*Ranges of accuracy of the applied model.*

Range, years	Observer 1, %	Observer 2, %	Mann [2], %
±10	27	28	33
±20	71	55	78

error were identified. These included individual interpretation of definitions, error associated with the application of the specifics of the definitions on sagittally sectioned skulls (see Materials and Methods), error in the measurement and recording of eight metrical attributes, and rounding error.

The investigators further educated themselves in the method of data collection, and re-collected the data on a smaller sample of maxillae, which excluded the females. The statistical analyses were performed, and two conclusions were apparent: first, the results by two observers were highly correlated, and second, there were no significant deviations between paired and ranked observations, which indicated that there was no significant interobserver error (Tables 2*a* and 2*b*). The results of this second replication indicate that the first facet of error can be minimized.

Despite the fact that the second facet of error is not quantifiable, it was estimated through achieving ranges of the percentage of accuracy which were similar to those of Mann [2], and does not reflect a failure of the model when applied to a different population (Table 3). However, following Lovejoy et al. [5], we note that the inaccuracy of the ages predicted by the method of Mann [2] is extreme (14.61 years), exceeding the inaccuracy of all other aging methods in Test 2 of Lovejoy et al., including the method of ectocranial suture closure (9.9 years). The bias (0.19 years) is similar to that of the revised pubic symphysis aging method (−0.4 years) [5].

The actual failure of the inverse prediction model arises both in the limits on quantifiable variability imposed by the method and in the biological variability, which cannot

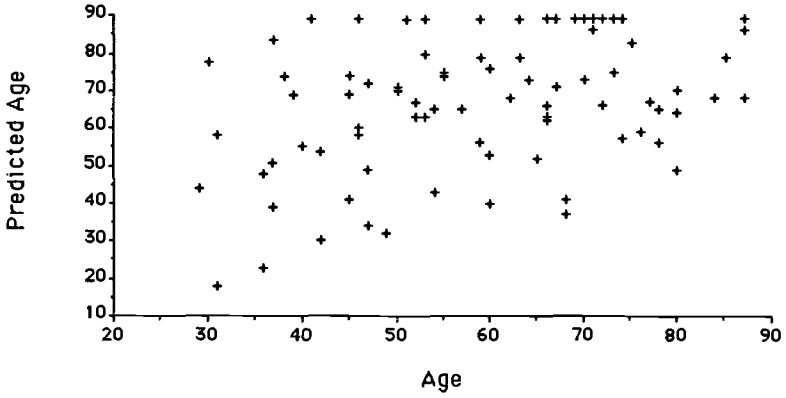

FIG. 1—*Observer 1, Test 2—verified age versus predicted age.*

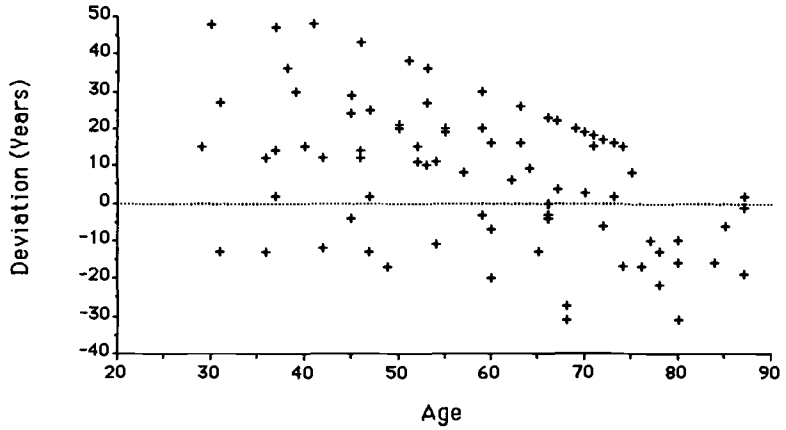

FIG. 2—*Observer 1, Test 2—verified age versus deviation in years.*

be accounted for in a linear estimation. The relationship between age and total suture score is not linear. Clearly, very young individuals will not have begun obliteration of maxillary sutures (total suture score = 0), while individuals beyond a certain age will exhibit complcted obliteration (total suture score = 4). The age progressive linear trend in suturc obliteration describes only a scgment within the entire trend. The inclusion of individuals of agcs less and greater than this age segment has the effect of decreasing the slope of the linear trend away from the correct slope. Suture obliteration may begin as early as 13 years [*1*] and continuc through the ninth decade. However, the upper limit of obliteration is not reached in Mann's sample and, in fact, the mean suture score in the eighth to tenth decade is less than that of the seventh. This results in a further decrease in the regression slope.

The graphs generated from the data of Mann [*2*] indicate that individuals between the ages of approximately 16 and 33 years can be aged within 15 years (Figs. 5 and 6). Beyond about age 33, variability in the total suture score rapidly increases, with the result that thc predicted ages show no dcmonstrable trend in accuracy when compared with the

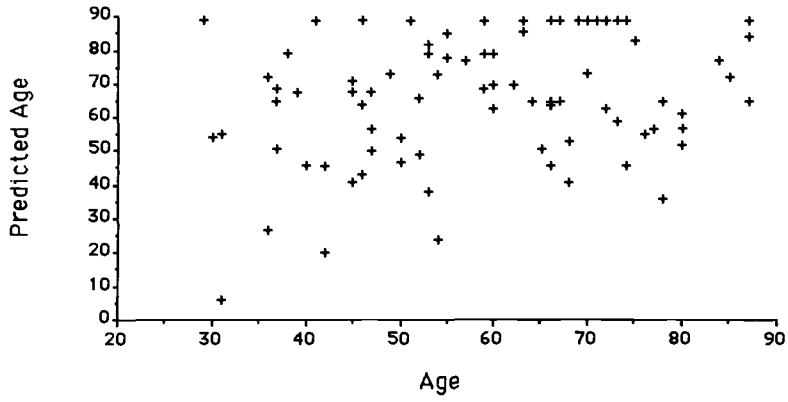

FIG. 3—*Observer 2, Test 2—verified age versus predicted age.*

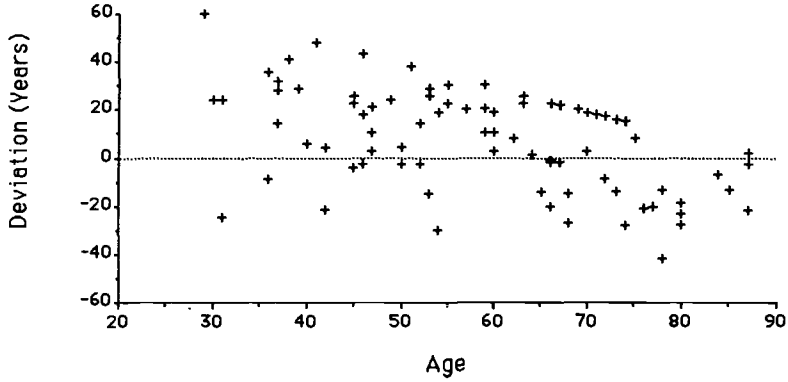

FIG. 4—*Observer 2, Test 2—verified age versus deviation in years.*

actual ages. It should be observed that one cluster of young individuals in Mann's sample gives greater correlation of the total suture score with age by virtue of a consistent low proportion of obliteration (Figs. 5 and 6). The graph of the age distributions of the samples shows the representation of young individuals in Mann's sample (Fig. 7). The apparent low correlation in the higher age group would not allow generation of a least squares linear regression with a slope significantly different from zero, if performed on this age group alone. In Mann's data [*2*], the high correlation of the total suture score with age in the young individuals masks a low correlation at higher ages and results in a significant linear regression.

Conclusions

Quantitative methods may be perceived as being more "scientific" and accurate than qualitative methods in the analysis of human skeletal material. In the case of this method, such a perception is dangerous. While the results of this study indicate that the method

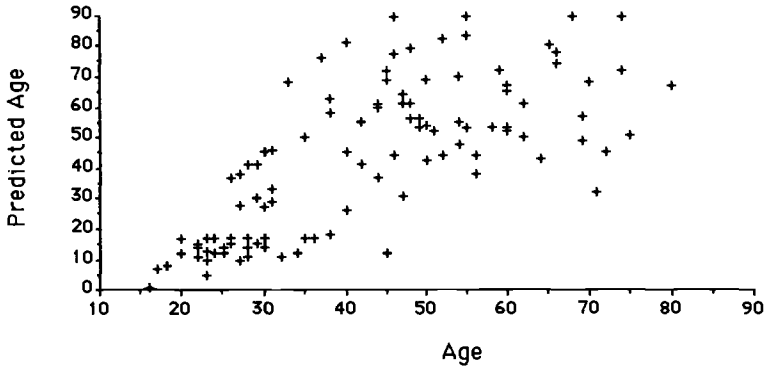

FIG. 5—*Mann* [2]—*age versus predicted age (males only).*

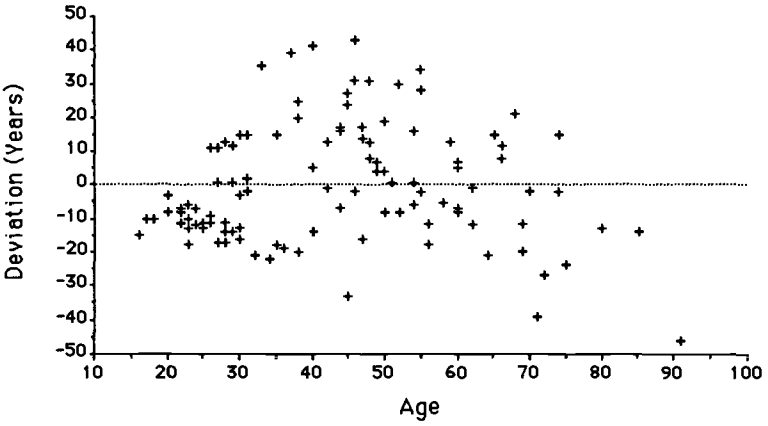

FIG. 6—*Mann* [2]—*age versus deviation in years (males only).*

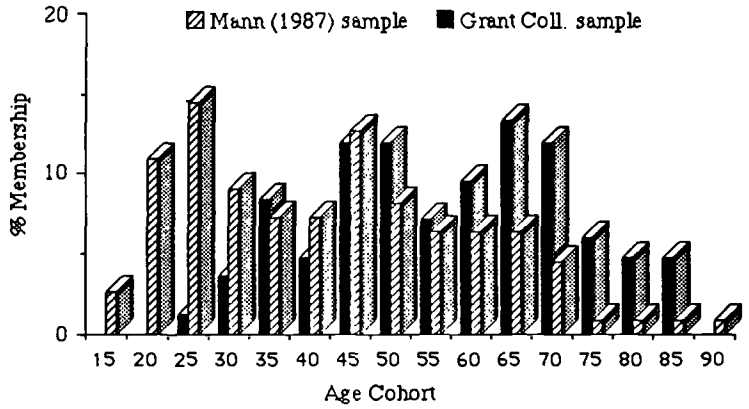

FIG. 7—*Percentage of membership by age cohort in the samples.*

is reproducible, it cannot corroborate the general statements regarding its usefulness put forward in previous publications [*1–3*]. It must be remembered that these results are derived only from a replicative study of males.

Acknowledgments

The authors would like to thank Robert Mann for his enthusiasm at the inception of this research and for the thesis copy which he provided. We also thank Dr. P. Stuart-Macadam (University of Toronto) for access to the Grant collection, Dr. J. Melbye (Erindale College, University of Toronto) for his comments, and Dr. S. G. Monckton for assistance with graphics. We are grateful to the review committee of this journal for their helpful comments.

References

[*1*] Mann, R. W., Symes, S. A., and Bass, W. M., "Maxillary Suture Obliteration: Aging the Human Skeleton Based on Intact or Fragmentary Maxilla," *Journal of Forensic Sciences*, Vol. 32, No. 1, Jan. 1987, pp. 148–157.

[*2*] Mann, R. W., *Maxillary Suture Obliteration: A Method for Estimating Skeletal Age*, unpublished Master's thesis, The University of Tennessee, Knoxville, TN, 1987.

[*3*] Mann, R. W. and Jantz, R. L., "Maxillary Suture Obliteration: A New Method of Estimating Age in the Human Skeleton," presented at the 40th Annual Meeting of the American Academy of Forensic Sciences, Philadelphia, PA, 15–20 Feb. 1988.

[*4*] Zar, J. H., *Biostatistical Analysis*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984, p. 156.

[*5*] Lovejoy, C. O., Meindl, R. S., Mensforth, R. P., and Barton, T. J., "Multifactorial Determination of Skeletal Age at Death: A Method and Blind Tests of Its Accuracy," *American Journal of Physical Anthropology*, Vol. 68, No. 1, Sept. 1985, pp. 1–14.

Address requests for reprints or additional information to
Kathy L. Gruspier
Department of Anthropology
University of Toronto
100 St. George St.
Toronto, Ontario
Canada M5S 1A1