J Forensic Sci, Jan. 1979, Vol. 24, No. 1

W. R. Maples, [1] *Ph.D. and P. M. Rice,* [1] *Ph.D.*

Some Difficulties in the Gustafson Dental Age Estimations

Estimation of the age at death of an individual from dental remains is commonly practiced. Ages of mineralization, crown completion, eruption, and root completion are usually accurate and reasonably easy to determine. After these and other growth processes are completed (at approximately age 30 years), age estimation becomes more difficult. Histological changes in bones and teeth have great potential for being the best age indicators for adult life.

Although Gustafson [*1,2*] has provided what seems to be the best operational method of dental age estimation to date, using semiquantitative (ordinal) indicators, his technique has had poor replicability. A number of investigators [*3-6*], using identical or modified techniques, have had difficulty in repeating the accuracy of his published results. To understand this difficulty, it is instructive to reexamine Gustafson's original publication [*1*] on this subject, which fortunately contains his raw data. It can be demonstrated that this paper contains a number of significant errors—technical, methodological, and interpretive—in the application of statistics.

Gustafson Statistical Analysis

After evaluating the accuracy of earlier impressionistic, nonquantitative methods of age determination, Gustafson developed a formula for coding degrees of age-change on six variables, such that the higher the coded score the greater the wear, and hence age, of the tooth. The summed scores ("points") of 19 teeth of known age were then fitted to a regression line. The legitimacy of the line drawn for this sample is questionable, however; most of the data points on the scattergram are loosely grouped at the high-point-high-age end, but the line drawn includes two points at the low-point-low-age end of the continuum, at some distance from the main grouping. This has the effect of creating strong linearity where none may exist, since the linear correlation coefficient recomputed with just the 16 items in the cluster is only +0.37.

Gustafson then apparently took the measurements from these 19 teeth, plus those from another 22 teeth, to develop and test a formal regression model for age estimation. Using a standard linear regression equation, he performed the regression of estimated age (as the dependent variable *Y*) on points, the summed indicators of age change. The results show a *Y* intercept of 11.43 years and a regression coefficient (or "slope") of 4.56.

It is at this point that an increasing number of errors enters into Gustafson's computations. Gustafson commendably attempts to assess the accuracy of his method, but he uses incorrect measures to do so. We shall return to this matter in a moment. More fundamental, however, is the fact that the same 19 teeth initially used to develop the regression model were among the 41 teeth that were later used in formalizing the model, and then

Received for publication 21 March 1978; accepted for publication 2 May 1978.

[1]Curator and chairman, Department of Social Sciences, Florida State Museum, and assistant professor, Department of Anthropology, respectively; University of Florida, Gainesville.

the accuracy of the model was in effect "tested" or evaluated by these same 41 teeth. Such a practice, that of testing a hypothesis, formula, or model on the data used to develop it, is universally condemned as unsound. It is circular reasoning and tautological, and does not constitute a test at all, but a demonstration. Moreover, because this sort of spurious test of a formula is predisposed to demonstrate its effectiveness in explaining the relationship between variables, Gustafson was able to claim a significantly higher degree of accuracy than have later researchers.

Another basis for his claim of higher accuracy is that he was calculating the wrong statistic or statistics with which to measure accuracy. He was treating, in effect, the potential for two kinds of error in his method or, more strictly speaking, two kinds of precision: (1) the precision or repeatability of his measurements (treated by what he calls "double determination" or "error of examination"); and (2) the precision of the method itself—the spread or dispersion of points around the regression line, that is, exactly how age can be predicted by the point scores.

With respect to the first kind of error, he shows that the coding process itself has a respectable degree of replicability. He gives the mean difference (measurement error) between successive codings of 37 teeth as 0.27 ± 0.17 points. Our recalculation suggests that the mean error is 0.76 points.

To determine "accuracy" (more correctly termed "precision," or the dispersion of values around the regression line), he calculates "the average deviation of a value from the regression line," or the "error of estimation" as $\sigma_{sy} - Y = \pm 3.63$ years. It is not immediately clear to us how he is calculating this. The standard procedure is to calculate the "error of estimation" as $Y - Y'$ (the true age minus the predicted value at the regression line). Such calculations give some positive deviations and some negative deviations because points lie both above and below the line. Since the regression line is a best-fit line minimizing distance between all points and the line, however, the positive and negative values cancel each other out. Hence, the sum of these deviations always is equal to zero and, thus, there is no way to compute an average of them. What Gustafson apparently did, as in his evaluation of the results of age estimations on Series III teeth in his Table 3, was calculate the "mean deviation," the average of the absolute value of the deviations, ignoring sign. True, this does give some indication of the degree of dispersion about the line, but this statistic is rarely used in modern statistical operations; more favored is the standard deviation, which allows further mathematical and inferential manipulations. Significantly, the mean deviation *cannot* be interpreted, as Gustafson appears to have done, as the standard deviation is interpreted, that is, in demarcating areas of probabilities of events in a normal distribution. Thus, his calculation of frequencies, or probabilities of cases falling within ±1, ±2, and ±3 mean deviation units from the real age [*1*, pp. 51-52], is meaningless.

Rather than calculating the mean deviation as an estimate of dispersive error, Gustafson should have calculated the standard error of the estimate, given by the following formula,

$$s_{y \cdot x} = (Y - Y')^2/(n - 2)$$

which is a measure of the variation of Y (age) around the regression line.

In any case, these errors are relatively meaningless in view of the fact that Gustafson derived his regression line (his Formula 2) from the wrong set of data. Gustafson's regression equation is

$$y = 4.56x + 11.43 \qquad r = 0.98 \tag{1}$$

The data used in developing this equation are published in his Table 5.

It appears that the regression formula developed by Gustafson was based on X as the point totals and Y as the estimated age as given in Table 5. Repeated recalculations of regression using these data gave a regression formula of

$$y = 4.46x + 11.56 \qquad r = 1.00 \tag{2}$$

These estimated age values, however, are the ones obtained by previous tests of his age-change indicators. Thus, he is doing a regression to find a best summarizing line for a scatter of points that is already a straight line.

The proper procedure for developing the regression line would have been to develop it from the age-change points and the known or real ages. Our calculated regression formula based on these values is

$$y = 4.26x + 13.45 \qquad r = 0.912 \tag{3}$$

Calculating the degree of dispersion about the line by means of the formula for the standard error of the estimate, we have found the error to be ± 7.03 years.

This still does not constitute a test of the regression model; a test would have to be done on independent data. However, it is theoretically possible, by using this error, to construct confidence intervals around the regression line to indicate the probability of accuracy of the estimate in expressing the true relationship between the two variables. As Blalock [7] explains, however,

> Since the quantity we are now estimating is no longer a single value but rather an entire line, our "interval estimate" will no longer be an interval but instead will consist of a band on either side of the least-squares line. At first thought, one might expect such a band to consist of two lines running parallel to the least-squares line. But such a band would imply that we knew the correct slope and that the only source of error came in estimating α. We must remember that there are now two quantities being estimated (α and β) and therefore two sources of error.

The confidence intervals around regression lines actually form two parabolic curves, coming closest to the line in the center and moving away from the line at the ends. Because an interval has to be computed for each point along the line, computations are tedious and rarely used [7].

Gustafson interprets his average deviation of ± 3.63 as standard deviation units of area under a normal curve whose mean is the regression line. According to this, two thirds of the cases are predicted between ± 3.6 years of accuracy by the equation; 95% are predicted between ± 7.3 years, and so on. This is not correct; as shown above, his figure of ± 3.63 years as the average deviation is not the correct statistic with which to do this, and it was obtained on the wrong set of data. The proper statistic would have been the standard error of the estimate which, on the correct data set to summarize the relationship between age and dental change, is ± 7.03 years. But in any case, because confidence intervals with regression equations are bounded by curves concave to the line rather than parallel lines on either side of it, the distribution of probabilities for predictive purposes cannot be so simply computed or interpreted.

Tests on Other Samples

Whenever the Gustafson regression has been used by other researchers to calculate the ages of teeth, the precision is less than that claimed by Gustafson. Although this has been particularly true when the published regression line is used (Eq 1), it would also be true when our corrected line for Gustafson's sample (Eq 3) is used.

For example, the standard error of the estimate $s_{y \cdot x}$ of the ages determined by Eq 3 applied to the control sample ($n = 51$) originally used by Burns and Maples [6] is ± 11.28 years, compared to the ± 7.03-year $s_{y \cdot x}$ derived on the original Gustafson sample. The $s_{y \cdot x}$ is usually higher when a regression is tested on an independent sample, but the difference encountered in this case is rather great. This might be explained by inconsistencies between individual observers, differences in populations from which the samples were taken, differences in sample selection procedures, or several other less likely causes.

Reconsideration of the Burns and Maples data [6] leads us to suggest that differences in the procedures or criteria for sample selection may be the primary causes for differences in precision. We say this for two reasons.

First, the Burns and Maples sample ($n = 219$ with complete data) was obtained from a dental clinic visited by mostly lower income patients, and many of the teeth displayed damage from caries. Gustafson [2] acknowledged that all of the six age changes scored by his technique are related to pathology; it is merely a question of degree when teeth are rejected or selected because of presence or absence of pathology. He admits that "when dealing with teeth from mouths that have been less well cared for, it is possible that the estimation will be too high" [1]. The Burns and Maples population, and therefore the sample too, probably contained more carious teeth than did the sample of Gustafson. Presumably, his sample was selected to exclude caries as much as possible to avoid errors in estimation stemming from pathology.

Second, when the regression-line-derived formula ($y = 5.6x + 4.66$) from the Gustafson scores of the Burns and Maples working sample ($n = 168$) was used on Gustafson's sample ($n = 41$) and on the Burns and Maples control sample ($n = 51$), the $s_{y \cdot x}$ values were ± 8.38 years and 10.59 years, respectively. It seems, therefore, that the use of a larger, more heterogeneous sample to derive a regression line yields more precise results when that line is tested on a more homogeneous and smaller sample than does the use of a line derived from a small homogeneous sample such as Gustafson's.

Thus, the Gustafson technique has been found to be lacking in precision when it is tested on other samples for the statistical reasons described earlier and also because of the sample Gustafson selected. The Burns and Maples data seem to suggest that careful selection of a very homogeneous working sample might produce a regression that loses precision when applied to the real world of variation.

Conclusion

The Gustafson technique was a significant contribution to forensic identification. Unfortunately, many statistical errors were present in the published articles. It now seems clear that greater precision can be achieved through modification of the original technique. One of the authors (W. R. M.) will report soon on multiple regression formulae that offer both increased precision and technical simplicity.

References

[1] Gustafson, G., "Age Determinations of Teeth," *Journal of the American Dental Association*, Vol. 41, No. 1, 1950, pp. 45–54.

[2] Gustafson, G., *Forensic Odontology*, American Elsevier Publishing Co., New York, 1966.

[3] Nalbandian, J. and Sognnaes, R. F., "Structural Age Changes in Human Teeth," in *Ageing: Some Social and Biological Aspects*, N. W. Shock, Ed., American Association for the Advancement of Science Publication No. 65, Chicago, 1960.

[4] Miles, A. E. W., "The Dentition in the Assessment of Individual Age in Skeletal Material," in *Dental Anthropology*, D. R. Brothwell, Ed., Symposia of the Society for the Study of Human Biology, Vol. V, Macmillan Co., New York, 1963, pp. 191–208.

[5] Johnson, C. C., "Transparent Dentine in Age Estimation," *Oral Surgery, Oral Medicine and Oral Pathology*, Vol. 25, No. 6, 1968, pp. 834–838.

[6] Burns, K. R. and Maples, W. R., "Estimation of Age from Individual Adult Teeth," *Journal of Forensic Sciences,* Vol. 21, No. 2, 1976, pp. 343-356.
[7] Blalock, H. M., *Social Statistics,* McGraw-Hill Book Co., New York, 1960.

Address requests for reprints or additional information to
W. R. Maples, Ph.D.
Department of Social Sciences
Florida State Museum
University of Florida
Gainesville, Fla. 32611