CLUSTERING ON THE BASIS OF LONGITUDINAL DATA

Emet D. Schneiderman,* Stephen M. Willis* and Charles J. Kowalski†

* Department of Oral and Maxillofacial Surgery, Baylor College of Dentistry, 3302 Gaston Avenue, Dallas, TX 75246, U.S.A.; and † Department of Oral Biology, and Center for Statistical Consultation and Research, The University of Michigan, Ann Arbor, MI 48109, U.S.A.

Abstract—A menu-drive PC program, ZDIST, for computing the distances between the estimated polynomial growth curves of subjects who have been followed longitudinally is described, illustrated, and made available to interested readers. These distances can be computed on the basis of the individual growth curves themselves and/or from estimates of individuals' growth velocity and acceleration curves. The resulting distance matrices can be saved in ASCII format and subsequently imported into any clustering program which accepts this type of input, e.g. SYSTAT.

Longitudinal studies   Growth curves   Cluster analysis   PC program

INTRODUCTION

Cluster analysis is commonly used in biomedical research to group similar entities (subjects, patients, objects) together. Good general accounts are given in [1–8]. An extensive bibliography, highlighting applications in biological systematics (taxonomy), but including references to the use of cluster analysis in ecology and biogeography, medicine, the social sciences, the earth sciences, other sciences and technology, and the arts and humanities is provided in [5]. Additional references to applications in medicine, psychology, archaeology and anthropology, phytosociology, economics, market research, and linguistics are given in [8]. And the volume of the literature in this area is expanding rapidly: it was suggested in 1971 that articles were appearing at the rate of 1000 per year [9], and we have no reason to suspect a recent decline. There have been, however, relatively few applications in longitudinal studies of growth, i.e. applications in which a clustering algorithm was applied directly to the repeated measurements made on each individual. In [10], changes in cluster characteristics over (exactly) two time points were considered, but these clusters were not formed on the basis of the repeated measurements. Most of the work in this area has concentrated either on allometry, which may be described as the study of differences in shape associated with size, or on methods for "factoring out" size, which is generally considered a nuisance parameter in taxonomic contexts [5]. There has also been some work on longitudinal data analysis in the problem obverse to cluster analysis, discriminant function analysis, where we are *given G* groups and asked to derive a rule for allocating "new" individuals to one of the groups on the basis of his/her growth profile. Useful background information is given in [11]. See also [12–14].

We suggest that, despite the fact that but few attempts to cluster individuals on the basis of longitudinal data have been made, it would often be of interest to identify subsets of individuals that are "growing similarly". The apparent reason for this void is a lack of appropriate software. While we are developing programs that *are* appropriate for the analysis of longitudinal data [15–27], and while there is a large number of programs,

with a myriad of options for metrics and algorithms, for performing cluster analyses, no bridge connecting these resources has been built. The purpose of the present paper is to describe, illustrate and make available a user-friendly, menu-driven PC program which computes the distances between individuals' growth, growth velocity and/or growth acceleration curves and saves these in an ASCII file which can then be read into a program which will perform the cluster analysis. The basic idea behind the distances computed by us is due to Zerbe [28]. The individuals comprising the study sample need not be measured at each of the planned times of measurement. The program ZDIST is written in GAUSS386i, but users need not have purchased or installed GAUSS386i to use our program, which stands alone. Hardware requirements and details concerning obtaining a copy are provided in the Appendix.

We consider first a brief description of cluster analysis. We then outline Zerbe's approach to computing the distances between growth curves. Finally, the technique and the program are illustrated using a sample of $N = 11$ achondroplastic children whose head circumference was measured (with substantial missing data) monthly from birth to 1 year of age [29].

CLUSTER ANALYSIS

Clustering is the grouping of similar objects. There is no *one* method of cluster analysis: indeed, it has been suggested [30] that a given clustering method can be characterized in terms of whether it does or does not possess some 45 properties, so that the total number of possible methods is of the order of 2^{45}. We concentrate on the class of methods which are generally called *hierarchical, agglomerative* procedures. Our aim is to provide enough background to indicate possible applications of our program, and enough terminology to allow a meaningful description of our example. There is no suggestion that this class of procedures is in any way "optimal", nor that other methods cannot be used.

Agglomerative methods all begin with N clusters each containing a single element, a "proximity" (either distance or similarity) matrix for these N clusters consisting of a single element, and a measure of distance between two clusters when each cluster contains one or more elements which provides the basis for linking the clusters together to form new ones [31]. The first step is to join the two nearest objects into a single cluster, so that we have N-2 clusters containing one object each and one cluster with two elements. The second step is to fuse the two nearest of the N-1 clusters to form N-2 clusters. We continue in this manner until at the $(N$-1)st step we join the two remaining clusters into a single cluster. A graphical representation of this process is called a *dendrogram*, or simply a tree diagram.

Our program provides the distance matrix for the N cases (clusters with but a single element). The user must decide on which measure of distance between clusters having one or more elements to employ. Among the many available are *single linkage*, defined as the smallest distance between elements of the clusters; *complete linkage*, where the largest distance is used; and *the centroid method*, where the distance between the centroids of the clusters is used. He/she must also decide at which stage to stop the process of fusion (where to "cut" the tree). The possibilities will depend on the program utilized to accomplish the cluster analysis. We describe several available in SYSTAT later. We turn now to a description of how the distances between individual cases are computed.

DISTANCES BETWEEN GROWTH CURVES

We consider longitudinal data sets of the form

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1T} \\ x_{21} & x_{22} & \ldots & x_{2T} \\ \vdots & \vdots & \ldots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{NT} \end{bmatrix}, \tag{1}$$

where x_{ij} denotes the value of the measurement under consideration for individual $i(i = 1, 2, \ldots, N)$ at time $t_j(j = 1, 2, \ldots, T)$. In (1), some of the x_{ij} may be missing, but we assume that such missing data points are "missing at random", i.e. that occurrences of missing data are not related to the values of neighboring measurements [32].

Following the procedure detailed in [33], we fit a polynomial of degree D_i the data from the ith individual and let $D = \max \{D_i\}$. We then estimate the growth curve for individual i by

$$\hat{x}_i(t) = [1, t, t^2, \ldots, t^D]\hat{\tau}_i, \tag{2}$$

where $\hat{\tau}_i$ denotes the estimated polynomial regression coefficients

$$\hat{\tau}_i = (W_i' W_i)^{-1} W_i' x_i \tag{3}$$

augmented with $D-D_i$ zero elements. Thus while a particular individual may require but a $D_i = 2$ degree equation to adequately describe his/her data, if $D = 5$, the 3×1 vector computed in (3) is made to be 6×1 by adding $D - D_i = 3$ zero elements (a quadratic is a quintic with three zero coefficients). An alternative approach, e.g. [29], is to refit each individual to $D = \max \{D_i\}$. This is considered below. The reason that either augmentation with zeros or refitting to D_{max} is necessary is that the distance between two growth curves (defined below) involves the difference of the estimated vectors of regression coefficients and consequently they must be of the same dimension. In (3), W_i is the within-individual (time) design matrix specific to the ith individual. For the definition and examples, see [33].

We next define the distance between $\hat{x}_i(t)$ and $\hat{x}_k(t)$ over the interval $[a, b]$ by

$$d_{ik} = \left[\int_a^b [\hat{x}_i(t) - \hat{x}_k(t)]^2 \, dt \right]^{1/2}, \tag{4}$$

which can be computed as

$$d_{ik}^2 = (\hat{\tau}_i - \hat{\tau}_k)' C (\hat{\tau}_i - \hat{\tau}_k), \tag{5}$$

where C has the (m, n)th element

$$c_{m,n} = \frac{b^{m+n-1} - a^{m+n-1}}{m+n-1}. \tag{6}$$

For growth velocity, (2) is replaced by its element-by-element derivative

$$\hat{v}_i(t) = [0, 1, 2t, 3t^2, \ldots, Dt^{D-1}]\hat{\tau}_i \tag{7}$$

and for acceleration by

$$\hat{a}_i(t) = [0, 0, 2, 6t, \ldots, D(D-1)t^{D-2}]\hat{\tau}_i. \tag{8}$$

For velocities, the C matrix in (5) has the (m, n)th element $c_{m,n} = 0$ if $m = 1$ or $n = 1$ and

$$c_{m,n} = \frac{(m-1)(n-1)[b^{m+n-3} - a^{m+n-3}]}{m+n-3} \tag{9}$$

otherwise. For accelerations, $c_{m,n} = 0$ if $m \leq 2$ or $n \leq 2$ and

$$c_{m,n} = \frac{(m-1)(m-2)(n-1)(n-2)[b^{m+n-5} - a^{m+n-5}]}{m+n-5} \tag{10}$$

otherwise.

The user may compute and save any or all of the resulting distance matrices. Note that accelerations will be available only if the degree of the polynomial fit is at least two. Also note that the distances considered above are defined over a particular interval of time, $[a, b]$. The user makes the final choice of $[a, b]$: there is a need to balance the choice of $[a, b]$, best made on biological grounds, with problems associated with extrapolating one or more polynomials beyond their ranges of observation. This should be clear from the example considered later.

Table 1.

Individual	T	R^2	t	p
1	8	0.968	0.144	0.8911
2	11	0.987	8.465	0.0001
3	4	0.997	0.522	0.6937
4	6	0.997	5.190	0.0139
5	7	0.993	3.709	0.0207
6	5	0.993	1.812	0.2117
7	12	0.955	2.703	0.0243
8	7	0.958	0.446	0.6789
9	8	0.979	1.455	0.2053
10	5	0.997	7.569	0.0170
11	4	0.976	0.708	0.6078

THE PROGRAM

The program is entirely similar in structure to that described and illustrated in [33]. The difference is that the present program, ZDIST, also provides the distances between the growth, velocity, and acceleration curves. The program is invoked with the single command *gsruni zdist*. Following the procedure detailed in [33], we determine either a common D for all cases, or a series of degrees $D_1, D_2, \ldots, D_N$ for each individual. In the former case we simply proceed with Zerbe's procedure; in the latter we determine $D = \max\{D_i\}$, and the program augments the vectors of coefficients with zeros as needed. Finally, the user is asked to enter the value of $[a, b]$ to be used in the rest of the analysis.

Once a common D and $[a, b]$ have been determined, we next produce the distances between the growth curves for each pair of individuals and print these and allow the user to save them. The user can also save the distances between the growth velocities and the accelerations. The numerical output is also saved in a file called ZDIST.OUT which can be modified, highlighted, etc., using a word processor and subsequently printed.

AN EXAMPLE

Our example is based on the data considered in [29, 33], consisting of head circumference measurements on $N = 11$ achondroplastic infants at $T = 13$ times of measurement (age in months from 0 to 12). Following the procedure outlined in [33], we try the SINGLE value of $D = 2$, and obtain plots of the fitted curves and the statistics (Table 1).

T is the number of (non-missing) observations; R^2 the square of the multiple correlation coefficient; and t is the value of the t-statistic, and p the corresponding p-value for the hypothesis H: $\tau_3 = 0$. Note that in our notation [15–27], τ_3 is the coefficient of the *quadratic* term, so that H is a test of whether this term may be dropped from the model. It is seen that τ_3 is significantly different from zero ($p < 0.05$) for 5 of the 11 cases and so, following [29], we choose to ACCEPT $D = 2$ for all 11 cases. We obtain the interindividual distance matrix as shown in Table 2.

Table 2. Interindividual distance matrix containing all distances between all 11 cases (C1–C11). Thus, the entry 6.36 is the distance between case 2 and case 1

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
C1	0										
C2	6.36	0									
C3	10.31	7.63	0								
C4	6.82	4.39	4.97	0							
C5	7.20	4.50	4.42	0.87	0						
C6	6.25	9.56	14.77	10.05	10.81	0					
C7	7.80	7.24	6.93	3.33	3.93	9.55	0				
C8	7.30	4.65	7.14	6.73	6.30	12.57	9.87	0			
C9	9.67	6.59	1.13	4.02	3.18	13.92	6.28	6.60	0		
C10	14.93	15.61	20.07	15.29	16.15	9.34	13.60	19.88	19.14	0	
C11	7.77	3.68	4.24	3.74	3.11	12.24	7.02	3.68	3.36	18.28	0

Single linkage method (nearest neighbor)

Tree diagram   Distances

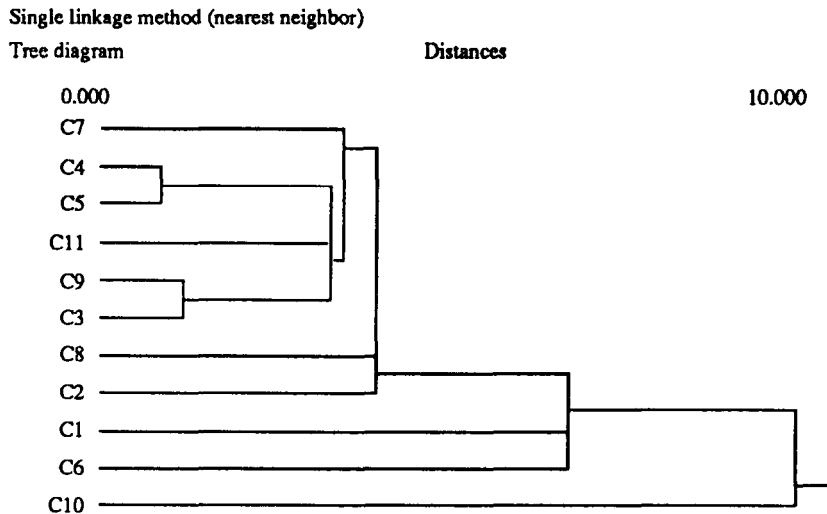0.000   10.000

Fig. 1.

Distance matrices for velocity and acceleration are also printed but not shown here. Data sets of this form (0s on the diagonal and only the lower elements of the matrix actually entered) may be imported, e.g. into the SYSTAT data editor and cluster analyzed. In SYSTAT, distance matrices are referred to as *dissimilarity* matrices and, while in the editor, the user must indicate that the data are of this type. Then using the JOIN function (the KMEANS option is not available for dissimilarity data) and single linkage (where the distance between clusters is the minimum distance between elements of the clusters), we obtain the dendrogram shown in Fig. 1.

DISCUSSION

It is seen from Fig. 1 that, at step 1, cases 4 and 5 are joined to form the first cluster (these have minimal distance $d_{45} = 0.87$). Then cases 3 and 9 are joined ($d_{39} = 1.13$). In the next steps we form two main clusters {C7, C4, C5, C11, C9, C3} and {C8, C2}. If we cut the tree at this point, the cases within these clusters are "growing similarly", while C1, C6 and C10 are "growing differently". If we wait until {C1, C6} is formed, we have identified four clusters, namely

$$\{C7, C4, C5, C11, C9, C3\}, \{C8, C2\}, \{C1, C6\}, \text{ and } \{C10\}.$$

We note that no matter where the tree is cut, C10 stands out as different from the rest. In this simple example, of course, this is obvious from an inspection of the distance matrix, but in more complicated problems the tree or dendrogram provides a useful visualization of the clustering process. It can also be seen that the clusters identified above do not reflect a simple separation of growth curves into those for which $D = 1$ and $D = 2$ are adequate fits. Referring to the p-values given earlier, $D = 2$ was indicated for C2, C4, C5, C7 and C10. While this *could* happen in a given situation, it will not be true in general that the composition of the clusters is dictated by the degree of the polynomial fit. The distance between a line and a quadratic can be small; the distance between two lines or two quadratics can be large.

It should be noted that the above results are for the fusion method of single linkage, where the distance between clusters is defined as the minimum distance between members of the two clusters. SYSTAT allows the user to select other linkage functions including complete (where the maximum distance is used), centroid, average and median. Other packages may offer different options.

While not considered explicitly as part of our example, the user may also perform the cluster analysis on the basis of distances between velocity and/or acceleration curves. Velocity (acceleration) curves determined as derivatives of estimated growth curves (as

is done in our program) tend to be reliable only over the time interval for which velocity (acceleration) could have been determined by fitting growth increments [16]. Thus, for example, Zerbe [28] took $a = 4.5$ and $b = 17.5$ for the velocity curves in a study where observations were made from 4 to 18 years at half-year intervals; and $a = 5$, $b = 17$ for accelerations. In our example, one might consider using $a = 1$, $b = 11$ for velocities; $a = 2$, $b = 10$ for accelerations. It will be recalled that the user of our program has control over the choice of a and b, and that plots are provided to facilitate this decision.

Finally, we consider one aspect of our example in more detail. We based the example on the data given in [29], and we followed them in fitting polynomials of degree $D = 2$ to each individual growth profile. It was on this basis that C10 was singled out as growing differently from the rest. Actually, this difference is not so much a function of the *data* for C10 as it is of the choice of $D = 2$ for this individual. Note that C10 was measured only at $t = 1$, 2, 3, 4, 6; and while a quadratic equation provided an extremely good fit ($R^2 = 0.997$) *over the interval of observation for this individual*, this equation extrapolates poorly (it turns down—a biologically improbable occurrence for a measurement such as head circumference) to the interval of interest [0, 12]. This is the reason for the large number of plots produced by our program; for allowing different individuals to have differing values of D; and for user-control over $[a, b]$. The plots identified the reason for the "aberrant" growth pattern for this individual *before* the distances were computed [33]; the choice of $D = 1$ is reasonable for C10 over [0, 12]; and the quadratic could have been used if the investigator was interested only in the growth for the first six months of life. We might have corrected this "error" before preparing our example, but thought the version presented above provided more justification for the structure of our program and is of heuristic value. It also demonstrates that if there *is* an aberrant grower in the sample, our program is likely to find him/her.

SUMMARY

A menu-driven PC program, ZDIST, for computing the distances between the estimated polynomial gowth curves of subjects who have been followed longitudinally is described, illustrated, and made available to interested readers. These distances can be computed on the basis of the individual growth curves themselves and/or from estimates of individuals' growth velocity and acceleration curves. The resulting distance matrices can be saved in ASCII format and subsequently imported into any clustering program which accepts this type of input, e.g. SYSTAT.

The method accommodates missing data: we assume that common times of measurement are *planned*, but allow some of the measurement sequences to be incomplete. The program features flexible plotting procedures which aid the user in determining the degree(s) of the polynomials to be fitted together with an interval over which the fitted curves are well-behaved and fairly represent the observed gowth profiles.

The method and program were illustrated using a data set for which common times of measurement were planned, but a substantial amount of missing data ensued. SYSTAT was used to cluster the matrix of distances between the growth curves and we showed that it was indeed possible to isolate subsets of similar growth curves on this basis.

REFERENCES

1. M. R. Anderberg, *Cluster Analysis for Applications*. Academic Press, New York (1973).
2. R. E. Blackith and R. A. Reyment, *Multivariate Morphometrics*. Academic Press, New York (1971).
3. M. G. Kendall, *Multivariate Analysis*. Griffin, London (1980).
4. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy*. Freeman, San Francisco (1963).
5. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*. Freeman, San Francisco (1973).
6. R. C. Tryon and D. E. Bailey, *Cluster Analysis*. McGraw-Hill, New York (1970).
7. J. van Ryzin, Ed., *Classification and Clustering*. Academic Press, New York (1977).
8. J. A. Hartigan, *Clustering Algorithms*. Wiley, New York (1975).
9. R. M. Cormack, A review of classification, *J. R. Stat. Soc. A* **134**, 321–367 (1971).
10. D. M. Dunn and J. M. Landwehr, Analyzing cluster effects over time, *J. Am. Stat. Assn* **75**, 8–15 (1980).

11. P. A. Lachenbruch, *Discriminant Analysis*. Hafner, New York (1975).
12. A. P. Azen and A. A. Afifi, Asymptotic and small-sample behavior of estimated Bayes rules for classifying time-dependent observations, *Biometrics* **28**, 989–998 (1972).
13. T. P. Burnaby, Growth-invariant discriminant functions and generalized distances, *Biometrics* **22**, 96–110 (1966).
14. K. Ulm, Classification on the basis of successive observations, *Biometrics* **40**, 1131–1136 (1984).
15. E. D. Schneiderman and C. J. Kowalski, Implementation of Rao's one-sample polynomial growth curve model using SAS, *Am. J. Phys. Anthropol.* **67**, 323–333 (1985).
16. E. D. Schneiderman and C. J. Kowalski, Implementation of Hills' growth curve analysis for unequal-time intervals using GAUSS, *Am. J. hum. Biol* **1**, 31–42 (1989).
17. E. D. Schneiderman, C. J. Kowalski and T. R. Ten Have, A GAUSS program for computing an index of tracking from longitudinal observations, *Am. J. hum. Biol.* **2**, 475–490 (1990).
18. T. R. Ten Have, C. J. Kowalski and E. D. Schneiderman, PC program for analyzing one-sample longitudinal data sets which satisfy the two-stage polynomial growth curve model, *Am. J. hum. Biol.* **3**, 269–279 (1991).
19. E. D. Schneiderman, S. M. Willis, T. R. Ten Have and C. J. Kowalski, Rao's polynomial growth curve model for une1ual-time intervals: a menu-driven GAUSS program, *Int. J. Biomed. Comput.* **29**, 235–244 (1991).
20. T. R. Ten Have, C. J. Kowalski and E. D. Schneiderman, A PC program for obtaining orthogonal polynomial regression coefficients for use in longitudinal data analysis, *Am. J. hum. Biol.* **4**, 403–416 (1992).
21. E. D. Schneiderman, C. J. Kowalski, T. R. Ten Have and S. M. Willis, Computation of Foulkes and Davis' nonparametric tracking index using GAUSS, *Am. J. hum. Biol.* **4**, 417–420 (1992).
22. E. D. Schneiderman, S. M. Willis, C. J. Kowalski and T. R. Ten Have, A PC program for comparing tracking indices in several independent groups, *Am. J. hum. Biol.* **4**, 399–401 (1992).
23. T. R. Ten Have, C. J. Kowalski, E. D. Schneiderman and S. M. Willis, Two SAS programs for performing multigroup longitudinal analyses, *Am. J. Phys. Anthropol.* **88**, 251–254 (1992).
24. C. J. Kowalski, Data analysis in craniofacial biology with special emphasis on longitudinal studies, *Cleft Palate-Craniofacial J.* **30**, 111–120 (1993).
25. C. J. Kowalski and E. D. Schneiderman, Tracking: concepts, methods and tools, *Hum. Evolution* (in press).
26. T. R. Ten Have, C. J. Kowalski, E. D. Schneiderman and S. M. Willis, A PC program for performing multigroup longitudinal comparisons using the Potthoff-Roy analysis and orthogonal polynomials, *Int. J. Biomed. Comput.* **30**, 103–112 (1992).
27. E. D. Schneiderman, S. M. Willis, C. J. Kowalski and T. R. Ten Have, A PC program for growth prediction in the context of Rao's polynomial growth curve model, *Comput. Biol. Med.* **22**, 181–188 (1992).
28. G. O. Zerbe, A new non-parametric technique for constructing percentiles and normal ranges for growth curves determined from longitudinal data, *Growth* **43**, 263–272 (1979).
29. D. V. Dawson, A. B. Todorov and R. C. Elston, Confidence bands for the growth of head circumference in achondroplastic children during the first year of life, *Am. J. Med. Genet.* **7**, 529–536 (1980).
30. I. J. Good, The botryology of botryology, *Classification and Clustering*, J. van Ryzin, Ed., pp 73–94. Academic Press, New York (1977).
31. G. A. F. Seber, *Multivariate Observations*. Wiley, New York (1984).
32. D. B. Rubin, Inference and missing data, *Biometrika* **63**, 581–592 (1976).
33. E. D. Schneiderman, S. M. Willis and C. J. Kowalski, PC program for estimating polynomial growth, velocity, and acceleration curves when subjects may have missing data, *Int. J. Biomed. Comput.* (in press).

About the Author—EMET DAN SCHNEIDERMAN received a B.A. and M.A. in anthropology from Northwestern University in 1978, and a Ph.D. in biological anthropology from The University of Michigan in 1985. While at The University of Mchigan he was affiliated with the Center for Human Growth and Development and began conducting research in the area of craniofacial growth. In collaboration with Joseph Mudar, Dr Schneiderman developed an integrated software system for the analysis of cephalometric radiographs (x-rays of the head). While on the orthodontics faculty of the University of Detroit School of Dentistry from 1985 to 1988, Dr Schneiderman created a computerized cephalometry laboratory. In 1988 Dr Schneiderinan went to the Baylor College of Dentistry in Dallas where he is the director of research for the Department of Oral and Maxillofacial Surgery and Pharmacological Sciences. Dr Schneiderman and co-investigator Dr Charles Kowalski have been funded by NIH/NIDR from 1988 to 1993 to conduct this biostatistical research from which this paper issued. Dr Schneiderman has more than 45 publications including two chapters and the monograph, *Facial Growth in the Rhesus Monkey*, published by Princeton University Press in 1992.

About the Author—STEPHEN M. WILLIS received the B.S. degree in mathematics from the University of Texas at Arlington in 1987. He has over 15 years of experience in clinical toxicology and is currently operations manager of a regional toxicology laboratory in Dallas. Mr Willis is also the lead programmer/systems analyst for the NIH/NIDR grant on longitudinal statistical methods with Drs Kowlaski and Schneiderman. He has played a major role in the development of user-friendly interfaces for programs that have broad applications in the biomedical sciences.

About the Author—CHARLES J. KOWALSKI received the B.S. in mathematics from Roosevelt University in Chicago in 1962, the M.S. in statistics from Michigan State University in 1965, and a

Ph.D. in biostatistics from The University of Michigan in 1968. Dr Kowalski then joined the faculty of the Department of Oral Biology at the University of Michigan School of Dentistry. Dr Kowalski served as the assistant director of the University's Statistical Research Laboratory from 1971 to 1978 and research scientist at the Dental Research Institute from 1978 to the present and directed the institute's biometrics laboratory. He has been full professor of dentistry and statistician at The University of Michigan since 1978. At various times Dr Kowalski has served as a consultant to the National Football League, Park, Davis and Co., Nymegen University, Lanchester Cleft Palate Clinic, the Department of Antiquities of the University of Alexandria in Egypt, the U.S. Veterans Administration and the Eastman Dental Center. Dr Kowalski has published more than 180 scientific papers, including numerous chapters and the book *A Mixed-Longitudinal Interdisciplinary Study of Growth and Development* published by Academic Press in 1979. Dr Kowalski's research has focused on the application of statistical methods to dental and oral research with special emphasis on measurement processes, their validity, reliability and calibration. Longitudinal data analysis and the computer implementation of polynomial growth curve models have also been and continue to be a major thrust of his research. Drs Kowalski and Schneiderman have been funded by NIH/NIDR from 1988 to 1993 to study and implement biostatistical methods for the analysis of longitudinal data in the form of user-friendly microcomputer programs.

APPENDIX. COMPUTER IMPLEMENTATION

A full set of PC programs for longitudinal data analysis, including this program, can be obtained on high density 5.25" or 3.5" diskettes (please request type) by sending $25 to defray the cost of handling and licensing fees. These programs require a 80386 or 80486 based personal computer (PC) running the MS-DOS operating system (version 5.0 or higher is recommended, although versions as low as 3.3 will suffice). 80386 computers *must* also be equipped with a 80387 math coprocessor. At least 4 mb of memory is required, and must be available to GAUSS386i, i.e. not in use by memory resident programs such as *Windows*. EGA or VGA graphic capabilities are required to display the color graphics; VGA or SVGA is suggested to display optimally the graphic results. Runtime modules are supplied with the programs so that no additional software (i.e. compiler or interpreter) is required to run these programs. One can create and edit ASCII data sets for use by these programs using the full screen editor supplied with MS-DOS version 5.0. The programs are written and compiled using GAUSS386i, version 3.0, require no additional installation or modification, and are run with a single command. When requesting the programs, address inquiries to the corresponding author and make checks payable to Baylor College of Dentistry.