



0010-4825(94)00049-2

## PC PROGRAM FOR ASSESSING THE RELATIONSHIP BETWEEN RATE OF CHANGE AND INITIAL VALUE

CHARLES J. KOWALSKI, EMET D. SCHNEIDERMAN and STEPHEN M. WILLIS  
 Department of Biologic and Materials Sciences, Dental School, and The Center for Statistical Consultation and Research (CJK), The University of Michigan, Ann Arbor, MI 48109, U.S.A.; and Department of Oral and Maxillofacial Surgery and Pharmacology (EDS and SMW), Baylor College of Dentistry, Dallas, TX 75266-0677, U.S.A.

(Received 7 January 1994; received in revised form 4 October 1994)

**Abstract**—A menu-driven PC program implementing Blomqvist's [*J. Am. Stat. Assn.* 72, 746-749, 1977] method for assessing the relationship between rate of change and initial value is described, illustrated and made available. It is shown that the naive approach to this problem—computing the correlation between the initial value and either the amount or rate of change—results in a negatively biased estimator. The extent of this bias can be dramatic and may lead investigators to conclude that a negative correlation is present when none exists; or that there is no correlation when in fact the correlation is positive. Blomqvist's (maximum likelihood) estimator avoids this bias, and is obtained by a transformation of the naive estimator.

Growth rates      Change      Initial value      Biased estimator      PC program

### INTRODUCTION

When analyzing longitudinal data, it is often of interest to study the relationship between the initial value of a measurement and its rate of change over time. One might ask, e.g. whether initially high/low values have differing prognostic significance. The naive approach to investigating this relationship would involve obtaining estimates of the slopes of the regression lines for each individual, say  $b_i$  for  $i = 1, 2, \dots, N$  and then correlating these with the intercepts, say  $m_i$ . This provides a (negatively) biased estimate of the relationship between change and initial value, and the extent of this bias can be dramatic. Specific, numerical examples appear in the literature [1-4] and these will be described in more detail (in the Discussion) after we have developed the theory behind an estimator which is free from this bias, and hence a basis for comparison. Suffice it to say here that the biased estimator can be significantly less than zero when there is no relationship between change and initial value, and it can not be significantly different from zero when a significant positive association exists. The source of this bias is most easily demonstrated in the simple case of two times of measurement in the context of the correlation between change,  $d_i = x_{i2} - x_{i1}$  and the initial value,  $x_{i1}$  (with just two time points, the slope is  $d_i$  and the intercept  $x_{i1}$ ).

Most demonstrations are based on scenarios which incorporate measurement error. Let  $X_{i1}$  and  $X_{i2}$  denote the *true* values for the  $i$ th individual at times 1 and 2. Because of the vagaries of measurement, we do not actually observe  $X_{i1}$  and  $X_{i2}$ , but rather, for individual  $i$ ,

$$x_1 = X_1 + \varepsilon_1$$

$$x_2 = X_2 + \varepsilon_2$$

where  $\varepsilon_1$  and  $\varepsilon_2$  are errors of measurement. If, as is usually done, we assume bivariate normality and that the errors of measurement are independent of the true scores [ $\text{cov}(\varepsilon, X) = 0$  for all combinations of subscripts on  $X$  and  $\varepsilon$ ] and of each other [ $\text{cov}(\varepsilon_1, \varepsilon_2) = 0$ ], then estimates of the association between *true* change and initial level

using *observed* change and initial level are negatively biased. This can be seen from the expression for the covariance  $x_1$  and  $x_2 - x_1$ , viz.,

$$\text{cov}(x_1, x_2 - x_1) = \text{cov}(X_1, X_2 - X_1) - \text{var}(\varepsilon_1). \quad (1)$$

Thus, the covariance of interest,  $\text{cov}(X_1, X_2 - X_1)$ , is estimated with negative bias when  $\text{cov}(x_1, x_2 - x_1)$  is used. This is because  $x_2 - x_1$  is not independent of  $\varepsilon_1$ . In fact,

$$\text{cov}(\varepsilon_1, x_2 - x_1) = \text{var}(\varepsilon_1)$$

the amount of the bias in (1). Thus we see that, when dealing with imperfect measurements,  $\text{cov}(x_1, x_2 - x_1)$  will underestimate  $\text{cov}(X_1, X_2 - X_1)$ . As noted earlier, one might infer a strong negative correlation when the true correlation is near zero; or one might conclude that there is no association when a relatively strong positive correlation exists. Examples of this phenomenon are given in [1-4] and are discussed in more detail later.

It is also true that it is not necessary to incorporate errors of measurement into the model to make the point that correlating change and initial value can produce misleading results. Supposing  $x_1$  and  $x_2$  are measured without error, Oldham [5] showed that the correlation between change and initial level is given by

$$\rho(x_1, x_2 - x_1) = \frac{\rho(x_1, x_2)\sigma_{x_2} - \sigma_{x_1}}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2 - 2\rho(x_1, x_2)\sigma_{x_1}\sigma_{x_2}}}. \quad (2)$$

It is seen that this correlation depends on  $\rho(x_1, x_2)$ ,  $\sigma_{x_1}$ , and  $\sigma_{x_2}$  and can vary, as these factors vary, to a dramatic extent. Consider, e.g. that if  $x_1$  and  $x_2$  are *independent* and if  $\sigma_{x_1} = \sigma_{x_2}$ , then  $\rho(x_1, x_2 - x_1) = -1/\sqrt{2} = -0.707$ . Thus even if  $x_1$  and  $x_2$  are selected from a table of *random* numbers, there will be a substantial negative correlation between change and initial value, which has erroneously lead some investigators to conclusions like, "Those with the most serious problem (high blood pressure or low scores on a stress test) showed the most improvement." In the case of just two times of measurement, when interest centers on the hypothesis that no change has occurred, Oldham [5] suggested that instead of studying  $x_1$  and  $x_2 - x_1$ , one should concentrate on  $D = x_2 - x_1$  and  $S = x_2 + x_1$  (or the mean,  $S/2$ ). These are related according to the expression

$$\rho(S, D) = \frac{\sigma_{x_2}^2 - \sigma_{x_1}^2}{\sqrt{(\sigma_{x_1}^2 + \sigma_{x_2}^2)^2 - (2\rho(x_1, x_2)\sigma_{x_1}\sigma_{x_2})^2}}.$$

It is seen that if no change has taken place (so, in particular,  $\sigma_{x_2}^2 = \sigma_{x_1}^2$ ),  $\rho(S, D) = 0$ . If in a given situation,  $\rho(S, D) \neq 0$ , this is a positive finding and it may be inferred that change did in fact occur. This approach is a simple special case of the use of *orthogonal polynomials*, an approach which is most useful in testing the no change hypothesis (for a more detailed discussion of orthogonal polynomials, and a PC program which may prove useful in like contexts, see [6]). It is still of interest to inquire as to whether change is associated with the initial value, and we consider this further in the following section.

## CHANGE AND INITIAL VALUE

Ragosa and Willett [3] suggested that there are four ways to approach the problem of assessing the relationship between change and initial value. The first, and least satisfactory, is the one considered above where the simple correlation  $r(x_1, x_2 - x_1)$  is used. The second involves "disattenuation" of  $r(x_1, x_2 - x_1)$  using an estimate of the reliability of the baseline measurement in (1). The third employs the correlation structure implied by a specified model for growth; and the fourth estimates the relation between change and initial value by modeling both individual growth and individual differences in growth. The latter approach is the one pursued here. The method was developed by Blomqvist [7], and applied in a study of blood pressure measurements in [4]. We employ their notation to facilitate comparison of the quantities computed by our program and the derivations and examples given in these publications.

BLOMQUIST'S PROCEDURE

Blomqvist's procedure is based on the model and quantities specified below. It will be seen that many of these quantities involve the following form of the "naive approach" mentioned earlier: the individual growth curves are fit by ordinary least squares (OLS) to obtain slopes (rates of change) and intercepts (initial values) for each subject. We then regress the slopes on the intercepts, obtaining an estimate of the regression coefficient of slope on initial value. This estimate is then corrected to avoid the bias illustrated earlier.

The expression

$$x_{ij} = m_i + b_i t_j + \varepsilon_{ij} \tag{3}$$

models the observation for individual  $i (i = 1, 2, \dots, N)$  at time  $t_j (j = 1, 2, \dots, T)$  with starting point  $t_1 = 0$ . The errors of measurement are assumed to be independently normally distributed with  $E(\varepsilon_{ij}) = 0$  and  $\text{var}(\varepsilon_{ij}) = \sigma_\varepsilon^2$ . We also assume

$$\begin{bmatrix} m_i \\ b_i \end{bmatrix} \sim BVN \left\{ \begin{bmatrix} \mu \\ \beta \end{bmatrix}, \begin{bmatrix} \sigma_m^2 & \sigma_{mb} \\ \sigma_{mb} & \sigma_b^2 \end{bmatrix} \right\},$$

i.e. that  $m_i$  and  $b_i$  have a bivariate normal distribution with respective means  $\mu$  and  $\beta$ , variances  $\sigma_m^2$  and  $\sigma_b^2$ , and covariance  $\sigma_{mb}$ . We let  $\theta = \sigma_{mb}/\sigma_m^2 =$  true regression coefficient of  $b_i$  on  $m_i$ , and  $\lambda = \sigma_\varepsilon^2/\sigma_m^2$ .

Blomqvist's procedure then proceeds as follows. For each of the  $N$  individuals, use OLS to estimate  $m_i$  and  $b_i$ . Compute and save the mean squared error (residual mean square) for each case. Then compute  $\hat{\theta}^*$ , the regression coefficient of slopes on intercepts. The maximum likelihood estimator (MLE),  $\hat{\theta}$ , is a simple adjustment of  $\hat{\theta}^*$ , viz

$$\hat{\theta} = \frac{\hat{\theta}^* + a_1 \hat{\lambda}}{1 - a_2 \hat{\lambda}}, \tag{4}$$

where

$$a_1 = \bar{t}/s_{tt}$$

$$a_2 = \sum t_j^2 / (T s_{tt})$$

and

$$\bar{t} = \frac{1}{T} \sum t_j, \quad s_{tt} = \sum (t_j - \bar{t})^2.$$

In (4),  $\hat{\lambda} = \hat{\sigma}_\varepsilon^2/s_m^2$ , where  $\hat{\sigma}_\varepsilon^2$  is the mean of the mean squared errors and  $s_m^2$  is the sample variance of the estimated intercepts.

Given  $\hat{\theta}$ , we can estimate the correlation between slopes and initial values (intercepts),  $\rho(b, m) = (\sigma_m/\sigma_b) \theta$  from  $r(b, m) = (s_m/\hat{\sigma}_b) \hat{\theta}$ , where  $\hat{\sigma}_b$  is the sample variance of the estimated slopes.

Wu *et al.* [4] show that

$$\lim_{N \rightarrow \infty} N \sigma_{\hat{\theta}}^2 = \left[ \left( \sigma_b^2 + \frac{\sigma_\varepsilon^2}{s_{tt}} \right) (\sigma_m^2 + a_2 \sigma_\varepsilon^2) - (\sigma_{mb} - a_1 \sigma_\varepsilon^2)^2 + \frac{(2T-2)(a_1 + a_2 \theta)^2 \sigma_\varepsilon^4}{(T-2)} \right] / \sigma_m^4$$

so that for large samples, letting *RHS* stand for the right-hand side of the above expression with estimates substituted for parameters, the approximate standard error of  $\hat{\theta}$  is  $SE(\hat{\theta}) = \sqrt{RHS}/N$  and an approximate (large sample) 95% confidence interval for  $\theta$  is

$$\hat{\theta} \pm 2SE(\hat{\theta}).$$

In addition to estimating  $\theta$ , output from the program can be used to: (i) compute a subject's expected rate of change; and (ii) the predicted value  $t$  time units from now. A subject's expected rate of change is given by

$$\hat{b}_i = \hat{\beta} + \hat{\theta}(m_i - \hat{\mu}), \tag{5}$$

where  $\hat{\beta}$  is the average of the estimated slopes for individuals. The predicted value  $t$  time units from now is

$$\hat{x}_{it} = m_i + \hat{b}_i t, \quad (6)$$

where  $\hat{\beta}_i$  is computed from (5). The utility of these quantities will be illustrated in the example considered below.

### THE PROGRAM

The program is written in GAUSS386i, but users need not have purchased nor installed GAUSS to run our program which stands alone. It is invoked by issuing the command *gsruni rbm* (this stands for the correlation or relationship ( $r$ ) between slope ( $b$ ) and intercept ( $m$ )). The user is prompted for the name and location of the (ASCII) file containing the values of the repeated measurements, and is asked to enter the times of measurement, say  $t_1, t_2, \dots, t_T$ . We, following [4], first transform these time points by subtracting  $t_1$  from each so that  $t_1=0$  is the first time of measurement (this makes the intercept equal to the initial value). We then compute and print estimates of the six parameters in the model, viz.,  $\hat{\mu}$ ,  $\hat{\beta}$ ,  $s_m^2$ ,  $\hat{\sigma}_b^2$ ,  $\hat{\sigma}_{mb}$ , and  $\hat{\sigma}_\varepsilon^2$ . We also print  $\hat{\lambda} = \hat{\sigma}_\varepsilon^2 / s_m^2$ , both  $r^*(b, m)$  and  $r(b, m)$ , and  $\hat{\theta}^*$  and  $\hat{\theta}$ .  $r^*(b, m)$  is the uncorrected correlation coefficient;  $r(b, m)$  the corrected version. Confidence intervals for  $\hat{\theta}$  and  $r(b, m)$  are also provided.

### AN EXAMPLE

Examples in which  $\hat{\theta}^*$  seriously underestimates  $\theta$  are plentiful [1–6]. We consider instead a data set which has been often used to illustrate longitudinal data analytic procedures in the literature, e.g. [8–12], for which the difference between  $\hat{\theta}^*$  and  $\hat{\theta}$  turns out to be less dramatic. It consists of the ramus heights of  $N=20$  boys at ages 8, 8.5, 9 and 9.5 years of age. These are reproduced in Table 1.

After specifying the name and location of this file, the user is asked to enter the times of measurement. Following [4], we transform these to 0, 0.5, 1, 1.5, so that  $t_1=0$  is the time of the baseline measurement. The program first uses OLS to fit lines to each of the 20 sets of measurements. The intercepts ( $m_i$ ), slopes ( $b_i$ ), the mean squared error (MSE)

Table 1. Ramus heights of 20 boys at different ages

Ind.	Age			
	8	8.5	9	9.5
1	47.8	48.8	49.0	49.7
2	46.4	47.3	47.7	48.4
3	46.3	46.8	47.8	48.5
4	45.1	45.3	46.1	47.2
5	47.6	48.5	48.9	49.3
6	52.5	53.2	53.3	53.7
7	51.2	53.0	54.3	54.5
8	49.8	50.0	50.3	52.7
9	48.1	50.8	52.3	54.4
10	45.0	47.0	47.3	48.3
11	51.2	51.4	51.8	51.9
12	48.5	49.2	53.0	55.5
13	52.1	52.8	53.7	55.0
14	48.2	48.9	49.3	49.8
15	49.6	50.4	51.2	51.8
16	50.7	51.7	52.7	53.3
17	47.2	47.7	48.4	49.5
18	53.3	54.6	55.1	55.3
19	46.2	47.5	48.1	48.4
20	46.3	47.6	51.3	51.8

Table 2. Intercepts, slopes, mean squared errors, and  $R^2$  values when lines are fit by OLS to ramus heights of 20 boys

Ind ( <i>i</i> )	$m_i$	$b_i$	MSE	$R^2$
1	47.94	1.18	0.053	0.942
2	46.49	1.28	0.021	0.980
3	46.21	1.52	0.021	0.986
4	44.86	1.42	0.104	0.924
5	47.75	1.10	0.037	0.953
6	52.62	0.74	0.032	0.916
7	51.57	2.24	0.329	0.905
8	49.35	1.80	0.705	0.742
9	48.34	4.08	0.126	0.988
10	45.37	2.04	0.269	0.906
11	51.20	0.50	0.008	0.954
12	47.83	4.96	0.889	0.945
13	51.96	1.92	0.046	0.980
14	48.27	1.04	0.009	0.987
15	49.64	1.48	0.006	0.996
16	50.78	1.76	0.024	0.988
17	47.06	1.52	0.046	0.969
18	53.60	1.30	0.157	0.870
19	46.47	1.44	0.129	0.909
20	46.22	4.04	0.864	0.922

and the value of  $R^2$  for each are computed and printed by our program, and shown in Table 2.

The estimates of the parameters in the model and the uncorrected estimates of the correlation and regression coefficients are then printed in the following form:

MUHAT = 48.667  
 BETAHAT = 1.868  
 VAR(M) = 6.494  
 VAR(B) = 1.352  
 COV(BM) = -0.655  
 VAR(E) = 0.194  
 LAMBDAHAT = 0.0303  
 RSTAR(BM) = -0.221  
 THETAHATSTAR = -0.101.

All of the above quantities would ordinarily be computed when following the naive approach described above. In particular, RSTAR and THETAHATSTAR refer to  $r^*(b, m)$  and  $\hat{\theta}^*$ , the uncorrected estimates and their confidence intervals:

Lower limit	Thetahat	Upper limit
-0.296	-0.085	0.127
Lower limit	R(BM)	Upper limit
-0.649	-0.186	0.278

The quantities with STAR in their name are the uncorrected, "naive" estimators of the correlation between slope and intercept; and of the regression coefficient for slope on intercept ( $\hat{\theta}^*$ ). It is seen that, in this example, the corrected values are quite close to the uncorrected values. The negative bias associated with the uncorrected estimators, while present, is relatively small for this data set.

To illustrate the use of (5) and (6), consider an individual with initial value  $m_i = 50$ . The expected rate of change for this subject is

$$\hat{b}_i = 1.868 - 0.090(50 - 48.667) = 1.748.$$

Noting that in our example rates are computed per half-year, this subject's predicted value  $t$  years from now is

$$\hat{x}_{it} = 50 + 2 * 1.748t = 50 + 3.496t.$$

For example, for this subject with initial value  $m_i = 50$  at 8 years of age, we predict  $\hat{x}_{it} = 53.496$  at 9 years of age.

## DISCUSSION

The example considered above, where the adjusted estimator differed but slightly from the unadjusted estimator, is not typical. This can be seen from equation (4) and those immediately following. In particular, it is seen that  $a_1$  and  $a_2$  depend on the design of the study. Their values depend on the number and spread of the times of observation. For a given design,  $\hat{\theta}$  and  $\hat{\theta}^*$  will be relatively close when  $\lambda = \sigma_t^2 / \sigma_m^2$  is small, i.e. when lines provide good fits to the data and the variation among intercepts is relatively large. Table 2 shows that both these conditions are satisfied in our example. As  $\lambda$  increases,  $\hat{\theta}$  will differ more from  $\hat{\theta}^*$ . Often the differences are quite dramatic, and these differences can lead to conflicting inferences [2, 5]. Examples exist where significant negative regression coefficients have been reported, the significance either vanishing or becoming significant in the positive direction when the adjustment is made. Thus, Feinleib *et al.* [13] demonstrated the negative association between initial value and rate of change in blood pressure found in the classic study by Jents [14] could be attributed entirely to the bias inherent in the naive estimator. Dwyer and Feinleib [1] gave another example involving blood pressure where the unadjusted and adjusted regression coefficients were, respectively,  $-0.28$  and  $-0.09$ . Blomqvist [7] reported a change from  $-0.22 \pm 0.032$  to  $0.16 \pm 0.064$ .

We might also mention that an estimator of the regression coefficient of slope on initial value which is somewhat easier to compute, and lies somewhere "in-between" the naive estimator and Blomqvist's estimator exists [13]. It does not require that the mean squared errors for the individual regressions on time be computed, but it retains some of the negative bias associated with the naive estimator, and is less efficient than that considered in this paper [4].

Finally, the reader will have noted that the analysis described above assumes a linear relationship between  $b$  and  $m$ . This will be true whenever, for each individual, the response variable is linearly related to time, and if individual values of the slope and intercept have a bivariate normal distribution. This will be a reasonable assumption in many biomedical research contexts, e.g. Hui and Burger [15, p. 753] justify fitting straight-line growth curves to epidemiologic data: "an important feature of many epidemiologic studies is that the follow-up of most of the individuals studied are sufficiently short that the response curves in the intervals can be approximated by straight lines." In the present context, where rate of change is the focus, there is also reason to expect a good approximation even when the true growth curves are quadratics [16]. In any case, a PC program for testing the linearity of the individual growth curves is available [12].

## SUMMARY

When analyzing longitudinal data, it is often of interest to study the relationship between the initial value of a measurement and its rate of change over time, e.g. to see whether initially high/low values have differing prognostic significance. A stand-alone, user-friendly, menu-driven PC program implementing Blomqvist's [7] method for assessing this relationship has been described, illustrated and made available to interested readers. It was shown that the naive approach to this problem—computing the correlation between the initial value and either the amount or rate of change—results in a negatively biased estimator. The extent of this bias can be dramatic and may lead investigators to conclude that a negative correlation is present when none exists; or that there is no correlation when in fact the correlation is significantly positive. Blomqvist's

maximum likelihood based estimator avoids this bias, and is obtained by a transformation of the naive estimator.

*Acknowledgement*—Supported by DE 08730 from the National Institute of Dental Research.

## REFERENCES

1. J. H. Dwyer and M. Feinleib, Introduction to statistical models for longitudinal observation. *Statistical Models for Longitudinal Studies of Health*, J. H. Dwyer, M. Feinleib, P. Lippert and H. Hoffmeister, Eds. Oxford University Press, Oxford (1992).
2. J. L. Fleiss, *The Design and Analysis of Clinical Experiments*. Wiley, New York (1986).
3. D. R. Ragosa and J. B. Willett, Understanding correlates of change by modeling individual differences in growth, *Psychometrika* **50**, 203–228 (1985).
4. M. Wu, J. H. Ware and M. Feinleib, On the relationship between blood pressure change and initial value, *J. Chron. Dis.* **33**, 637–644 (1980).
5. P. D. Oldham, A note on the analysis of repeated measurements of the same subjects, *J. Chron. Dis.* **15**, 969–977 (1962).
6. T. R. Ten Have, C. J. Kowalski and E. D. Schneiderman, PC program for obtaining orthogonal polynomial regression coefficients for use in longitudinal data analysis, *Am. J. Hum. Biol.* **4**, 403–416 (1992).
7. N. Blomqvist, On the relation between change and initial value, *J. Am. Stat. Assn* **72**, 746–749 (1977).
8. R. C. Elston and J. E. Grizzle, Estimation of time-response curves and their confidence bands, *Biometrics* **18**, 148–159 (1962).
9. C. R. Rao, The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika* **52**, 447–458 (1965).
10. J. E. Grizzle and D. M. Allen, Analysis of growth and response curves, *Biometrics* **25**, 307–318 (1969).
11. H. Goldstein, *The Design and Analysis of Longitudinal Studies*. Academic Press, New York (1979).
12. 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).
13. M. Feinleib, M. Halperin and R. J. Garrison, Relationship between blood pressure and age: regression analysis of longitudinal data, Presented at the 97th Annual Meeting of the American Public Health Association, Philadelphia (1969).
14. R. M. Jense, Age variation of systolic blood pressure in United States Army Officers, *Am. J. Hyg.* **20**, 574–603 (1934).
15. S. L. Hui and J. O. Burger, Empirical Bayes estimation of rates in longitudinal studies, *J. Am. Stat. Assn* **78**, 753–760 (1983).
16. D. G. Seigel, Several approaches for measuring average rates of change for a second degree polynomial, *Am. Stat.* **29**, 36–37 (1975).

**About the Author**—EMET D. SCHNEIDERMAN received a B.A. and M.A. in anthropology from Northwestern University in 1978, and the Ph.D. in biological anthropology from The University of Michigan in 1985. While at The University of Michigan 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, 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 Schneiderman went to the Baylor College of Dentistry in Dallas where he is associate professor and director of research for the department of oral and maxillofacial surgery and pharmacology. Dr Schneiderman played a major role in initiating a new Ph.D. program in craniofacial biology at Baylor in 1993. Dr Schneiderman and co-investigator Dr Charles Kowalski have been funded by NIH/NIDR from 1988 to 1994 to conduct the biostatistical research from which this paper issued. Dr Schneiderman has more than 70 publications including two chapters and the monograph, *Facial Growth in the Rhesus Monkey*, published by Princeton University Press in 1992.

**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 the 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, Lancaster Cleft Palate Clinic, the Department of Antiquities of the University of Alexandria in Egypt, the US Veterans Administration and the Eastman Dental Center. Dr Kowalski has published more than 200 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 1994 to study and implement biostatistical methods for the analysis of longitudinal data in the form of user-friendly microcomputer programs.

**About the Author**—STEPHEN M. WILLIS received the B.S. degree in mathematics from the University of Texas at Arlington in 1987. Mr Willis 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 Kowalski and Schneiderman. Mr Willis has played a major role in the development of user-friendly interfaces for programs that have broad applications in the biomedical sciences. He has coauthored more than 20 scientific publications concerned with these programs. Mr Willis is also an accomplished amateur astronomer.

## APPENDIX

A set of PC programs, including this and related procedures can be obtained on 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.