British Journal of Orthodontics/Vol. 24/1997/61–66

Multiple Linear Regression as an Analytical Tool in Cephalometric Studies

JOS M. H. DIBBETS, PH.D.

Department of Orthodontics, Philipps-University, Marburg, Germany

CARROLL-ANN TROTMAN,* B.D.S., M.A.

JAMES A. McNAMARA, JR,*,** D.D.S., PH.D.

Departments of *Orthodontics and Pediatric Dentistry, and **Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan, USA

LEO TH. VAN DER WEELE

Division of Applications, Computer Centre, University of Groningen, Netherlands

JANINE E. JANOSKY, PH.D.

Department of Family Medicine and Clinical Epidemiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Abstract. *When the effect is studied of a factor like 'orthodontic therapy' on linear craniofacial growth, the concomitant consequence of age and gender on size cannot be ignored. The methodologically correct solution is division of the study group into smaller units, each of which is homogeneous with respect to age, gender, and therapy, and to compare these with matched controls. Yet, apart from matched controls being hard to find, this method of subdivision has the serious drawback that smaller groups decrease statistical power.*

A solution without the need to create sub-groups lies in the application of multiple linear regression analysis. It has been applied to biological data in other studies, but verification of the outcome has not been reported so far. Indeed, testing the mathematical assumptions underlying the regression model created unresolvable obstacles and, therefore, it was decided to perform verification by means of practical examples. Two separate tests for the applicability of the multiple linear regression method, on different data, with differing predictor sets, and with different control samples have been performed.

Index words: Cephalometry, Generalized Linear Models, Multiple Linear Regression, Multivariate Analysis, Variance Reduction.

Refereed Paper

Introduction

Linear craniofacial measures in children vary because of age, gender, other factors like orthodontic therapy or hormone therapy, and simply because some children are big while others are small. Size variance, therefore, complicates the comparison of linear dimensions between children. The preferred method to reduce size variance is to create sub-groups of children selected on age, gender, and therapy, and to compare these with matched controls. The disadvantage of sub-groups, however, is the inherent erosion of sample size and, therefore, loss of statistical power. For example, to separate males from females the sample requires a division in two parts; age may require a further division of each gender group in arbitrary age classes and the factor being studied, e.g. orthodontic therapy, finally, makes that the original study group at least has to be reduced in 2 (gender) × 2 (age) × 2 (therapy) = subgroups, each of these containing (on average) only one-eighth of the original number of individuals. It is obvious that few studies will survive this rigorous, however necessary, method of subdivision.

A solution which eliminates part of the confounding size variance in biological data without the need for creating sub-groups lies in the application of mathematical statistical techniques from the family of Generalized Linear Models (Landauer, 1962; Solow, 1966; Harris, 1971; Howells, 1971; Kowalski, 1972; Cleal *et al.*, 1979; Hallers ten-Tjabbes, 1979; Bookstein, 1982; Vark van, 1985; Cheverud and Richtsmeyer, 1986; Finkelstein *et al.*, 1988; Buschang *et al.*, 1990; Schneiderman *et al.*, 1993; Schneiderman and Kowalski, 1994). One of these statistical techniques is multiple linear regression (Draper and Smith, 1966; Smillie, 1966). A linear regression model assesses a linear relationship between one dependent and one or more predictors: e.g. size (dependent) and age (predictor)—larger in children typically being associated with them being older. Due to random effects and individual differences in size and maturation, the collected data points will be scattered around the computed least squares regression line.

Multiple linear regression analysis allows several pre-

0301-228X/97/001000+00$02.00

dictors to be put in the model simultaneously. In a multidimensional space, each of them is given a 'weight'. As the predictors are weighted, instead of being used to create subgroups, the resulting coefficients represent predictor dependent estimates rather than group averages. The application of the technique assumes a linear relationship between the dependent and predictor variables. The data itself need not to be normally distributed, as is often wrongly assumed (Smillie, 1966). They may be qualitative as well as quantitative or even can be a dichotomy like gender (Cox and Wermuth, 1992; Lauritzen and Wermuth, 1989). The error term or residual is considered to be independent, normally distributed with a mean of zero and a constant variance. The application is relatively robust and the results are generally accepted without further verification of the model (Harris, 1963; Wells, 1970; Keeling *et al.*, 1989; Meyers, 1992).

One way to verify the outcome of a regression analysis is to test if the assumptions for the residuals have been violated. We seriously tried to do so, but eventually had to drop the attempt because unresolvable obstacles made it impossible. Therefore, it was decided to test the model by means of practical examples, i.e. to compare the estimated coefficients with real values. To achieve that goal, those predictors had to be selected for which associations with cephalometric measures could be estimated by regression analysis, as well as computed from the original data. Obviously, these predictors are age and gender. Thus, the association with age and gender for several craniofacial measures was estimated by regression analysis, and then compared to gender difference and age increase per year computed directly from the data. Comparison of the outcome from these procedures will reveal if the regression model fitted the data adequately. This demonstration does not constitute and cannot replace testing of the mathematical assumptions underlying the statistical model, but it does provide strong evidence in support of the method.

Two separate tests on different data, with differing predictor sets and with different control samples will be performed.

Subjects and Methods

The present data were taken from three sources:
1. The Groningen, Netherlands, longitudinal elementary school growth study referred to as the Groningen Study (de Bruin, 1993).
2. The study of indications for tonsillectomy and adenoidectomy at the Children's Hospital of Pittsburgh USA, referred to as the Pittsburgh Study (Paradise and Bluestone, 1976);
3. The Ann Arbor USA longitudinal elementary school growth study, referred to as the Michigan Study (Riolo *et al.*, 1974).

In the Groningen Study 181 'average schoolchildren' were documented longitudinally with yearly cephalograms between the ages 6–18 years. This sample was divided in 11 age classes, the youngest from 6.5 to 7.5 years of age and the oldest from 16.5 to 17.5 years of age. For each age class at least 16 cephalograms were selected randomly, such that in the end each person appeared only once. This set may be considered a cross-sectional selection of Groningen children between 6·5 and 17·5 years of age. Roentgenographic enlargement was corrected. In the Pittsburgh Study cephalograms of 207 enrolled Caucasian children aged 3–13 years who had not (yet) undergone tonsillectomy or adenoidectomy were selected from a much larger group on the basis of the technical adequacy. Only one cephalogram per child was used and there has been no correction for roentgenographic enlargement (estimated at 11 per cent). In the Michigan Atlas 83 individuals who were documented longitudinally are depicted with tables and graphs for boys and girls separately between the ages 6–16 years. The roentgenographic enlargement of almost 13 per cent was not compensated.

Although the results of the analysis of airway parameters in the Pittsburgh Study constitute a separate study, a brief description is necessary because they will be introduced to increase the number of predictors in the regression model. For each child, observations were made concerning the degree of lip posture when the child was distracted or in repose, and rated on a 4-point scale: 1 = none, 2 = mild, 3 = moderate, and 4 = severe opening. The nasopharyngeal depth and horizontal pharyngeal tonsil shadow were measured on the lateral cephalogram. Nasopharyngeal depth was determined by measuring the shortest distance from the nasopharyngeal surface of the soft palate to the closest point on the posterior pharyngeal wall. Statistics of the independent variables for the Pittsburgh Study population are displayed in Table 1. Their associations were assessed by Pearson or Spearman correlation coefficients and are depicted in Tables 2, 2.1 (females), and 2.2 (males).

The method is a straightforward multiple linear regression model $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p + S$ (Smillie, 1966). Factors like age and gender were considered to be the 'X' or predictors; b_0 is a constant intercept; and b_1, b_2, . . ., b_p are beta-weighs corresponding to each of the predictors. S represents the error term or residual. The 'Y' or dependent variables consisted of conventional cephalometric linear measures. The choice of which cephalometric measures to use, will be explained at the end of this section. The software was from SYSTAT (1990, updated regression module).

In the Groningen study the average gender difference for the whole sample was computed along with the average yearly increase between 7 and 17 years of age. The Michigan Study is published as an atlas. It had to match the Pittsburg data and therefore age was limited to 13 years. The difference between boys and girls in each yearly age class from 6 to 13 years was calculated and then averaged, to obtain the mean gender difference. The yearly increase between 6 and 13 years in both boys and girls was calculated and averaged as non-gender specific growth per year.

As a first test, age and gender beta-weights for craniofacial measures in the Groningen Study were estimated by the regression model and then compared with the computed averages. As a second test, in the Pittsburgh study, the number of predictors was extended from two to five. It is not known which subset of all possible predictors constitutes a 'good' model, and therefore it had to be verified if the addition of predictors would influence the variance

explained by age and gender. Thus, besides age and gender, the three following predictors were added: lip posture, nasopharyngeal depth, and tonsil width. Beta-weights were estimated for all five predictors, but only the outcome for age and gender are reported and compared to the Michigan Study. The Michigan Study was introduced as an external reference with landmark definitions and tracing methodology generally being the same.

Interpretation of beta-weights

A beta-weight quantifies the change in a given cephalometric measure (as a dependent variable) for each step in the unit value of the predictor. The association between the dependent variable and the predictor therefore is proportional. For the predictor age the unit value is 'years', which means that a given cephalometric measure changes one beta-weight for each year of age difference. For the predictor gender, going from female to male, the unit value was 'one' which means that a cephalometric measure in boys was one beta-weight different from that in girls.

Considered were those linear cephalometric measures that had a significant gender and age coefficient in the regression analysis, and that had 'a match' in the Michigan Study. Redundancy was avoided. Separate regression models were fitted for each of the dependent cephalometric measures. The multiple *R*-square was taken as the proportion of total variation in the cephalometric measures accounted for by linear prediction. The level of significance was set at alpha = 0·05.

Results

Statistics of the predictors in the Pittsburg Study population are displayed in Table 1. Their mutual associations were assessed by Pearson or Spearman correlation coefficients and are depicted in Tables 2, 2.1, and 2,2. Even though some of the correlation coefficients were significantly different from zero, multicollinearity was slight because all of the coefficients were quite small.

The cephalometric measures are listed in Table 3.

Beta-weights for the Groningen Study and averages computed from the Groningen Study for gender and age are listed in Table 4.

Beta-weights for the Pittsburgh Study for two of the five predictors—namely, gender and age—and averages computed from the Michigan Study are listed in Table 5. The results for lip posture, sagittal airway and tonsil width are not relevant for the present tests, because they were introduced only as an expansion of the number of independent variables. It is sufficient to note that these three predictors produced many significant beta-weights.

TABLE 1 *Demographic and clinical variables for the Pittsburgh Study sample* (n = *207)*

Variable	Range	Mean	S.D.
Age (years)	3–13	7·1	2·0
Lip posture value	1–4	2·0	0·9
Nasopharyngeal depth (mm)	0·0–15·2	4·6	2·9
Tonsil width (mm)	5·6–22·9	13·9	3·4

Gender: female:male, 44:56 per cent.

TABLE 2 *Correlation coefficients between the independent variables in the Pittsburgh Study sample*

	Age	Lip posture	Nasopharyngeal depth
Lip posture	−0·03		
Nasopharyngeal depth	0·14*	−0·51*	
Tonsil width	0·06	0·24*	−0·13

* $P < 0.05$.

TABLE 2.1 *Correlation coefficients between the independent variables in the Pittsburgh Study sample, females only*

	Age	Lip posture	Nasopharyngeal depth
Lip posture	−0·01		
Nasopharyngeal depth	0·23*	−0·54*	
Tonsil width	0·06	0·21*	−0·25

* $P < 0.05$.

TABLE 2.2 *Correlation coefficients between the independent variables in the Pittsburgh Study sample, males only*

	Age	Lip posture	Nasopharyngeal depth
Lip posture	−0·05		
Nasopharyngeal depth	0·07	−0·47*	
Tonsil width	0·06	0·26*	−0·03

* $P < 0.05$.

TABLE 3 *Cephalometric measures compared in the combined three studies*

Sella–nasion	S–n
Sella–posterior nasal spine	S–PNS
Sella–articulare	S–Ar
Articulare–posterior nasal spine	Ar–PNS
Articulare–gnathion	Ar–Gn
Lower incisor edge–mandibular plane perpendicular	LIE–MPP
Gonion–pogonion	Go–Pg
Nasion–menton (total face height)	N–Me
Nasion–anterior nasal spine (upper face height)	N–ANS
Anterior nasal spine–menton (lower face height)	ANS–Me

Discussion

The Groningen Study (Table 4)

The variance explained by linear regression is given by 100 times the R-square, ranged from 31 to 67 per cent and may be considered high for biological data. Gender differences were accounted for by the regression model in amount as well as in direction; the beta estimates were between 0·1 and 0·7 millimetre, mean 0·3 mm, different from the corresponding computed averages. The cephalometric measures in boys were always larger than in girls. Growth increments per year, estimated beta-weights and computed averages, differed only by 0·1 mm per year. Thus, the beta-weights for the Groningen Study were fairly exact estimates of the sample parameters.

TABLE 4 *The Groningen Study*

Variable	Gender difference (male minus female)			Age (growth per year)			
	AvGr	BGr	(SE)	AvGr	BGr	(SE)	*R*-Square
S–N	1·3	1·5	(0·39)	0·5	0·6	(0·06)	0·39
S–PNS	0·8	1·0	(0·39)	0·8	0·7	(0·06)	0·45
S–Ar	0·6	0·8	(0·39)	0·8	0·7	(0·06)	0·44
Ar–PNS	0·5	0·6+	(0.34)	0·6	0·6	(0·05)	0·42
Ar–Gn	1·1	1·7	(0·68)	1·9	2·0	(0·11)	0·67
LIE–MPP	0·7	1·0	(0·37)	1·0	0·9	(0·06)	0·59
Go–Pg	0·6	1·0+	(0·57)	1·5	1·4	(0·09)	0·59
N–Me	1·8	2·5	(0·86)	1·9	1·9	(0·13)	0·54
N–ANS	1·0	1·3	(0·42)	0·9	0·9	(0·06)	0·53
ANS–Me	0·9	1·3+	(0·70)	1·0	0·9	(0·11)	0·31

AvGr: average gender difference or average growth per year in the Groningen Study sample ($n = 181$).
BGr: estimated significant ($P < 0.05$) beta-weights for gender and age ($+P < 0.1$).
(SE) standard error for the beta-weights.
Ideally, the computed average values AvGr should equal the estimated beta-weights BGr.
The data have been corrected for radiographic enlargement.
Cephalometric landmarks: see Table 3.

TABLE 5 *The Pittsburgh Study compared to the Michigan Study*

Variable	Gender difference (male minus female)			Age (growth per year)			
	AvMi	BPitt	(SE)	AvMi	BPitt	(SE)	*R*-Square
S–N	3·3	2·6	(0·43)	0·9	0·8	(0·11)	0·39
S–PNS	2·3	1·5	(0·37)	0·9	1·0	(0·09)	0·44
S–Ar	1·7	1·8	(0·38)	1·0	0·9	(0·10)	0·35
Ar–PNS	1·9	2·0	(0.37)	0·6	0·5	(0·09)	0·29
Ar–Gn	3·0	2·8	(0·65)	2·4	2·2	(0·16)	0·54
LIE–MPP	1·8	1·5	(0·45)	1·0	0·9	(0·14)	0·33
Go–Pg	1·2	1·5	(0·55)	1·9	1·7	(0·14)	0·48
N–Me	4·0	2·6	(0·68)	2·5	2·6	(0·17)	0·60
N–ANS	0·8	0·7	(0·39)	1·4	1·4	(0·10)	0·56
ANS–Me	3·4	1·7	(0·52)	1·1	1·0	(0·13)	0·43

AvMi: average gender difference or average growth per year in the Michigan Study sample ($n = 83$).
BPitt: estimated significant ($P < 0.05$) beta-weights for gender and age in the Pittsburgh Study sample ($n = 208$).
(SE) standard error for the beta-weights.
Ideally, the computed average values AvMi should equal the estimated beta-weights BPitt.
Radiographic enlargement (11–13 per cent) has not been corrected for.
Cephalometric landmarks: see Table 3.

The Pittsburgh Study compared to the Michigan Study (Table 5)

The variance explained was equally high as in the Groningen study and ranged from 29 to 60 per cent. Gender differences were accounted for by the regression model in amount as well as in direction, albeit that the Pittsburg estimates were between 0·1 and 1·7 mm, average 0·6 mm, different from the computed Michigan averages. Growth increments per year and estimated beta-weights differed on seven occasions by 0·1 mm per year and twice by 0·2 mm. Let us consider the effect of this 0·2 mm discrepancy. The average growth reported in the Michigan Study for one of these two, the dimension articulare–gnathion, over a period from 6 to 13 years of age is 17 mm. The regression model underestimated this amount of growth by $7 \times 0{\cdot}2 = 1{\cdot}4$ mm. All in all, the beta-weights for the Pittsburgh Study may be considered acceptable estimates when compared to the Michigan sample parameters.

Comparison of the Studies

Standard Error. The standard error for the estimated Beta weights was almost equal for the Groningen study compared to the Pittsburgh study.

Gender. Gender differences in the Groningen Study, beta-weights as well as computed average values, are

remarkably smaller than in both the Pittsburgh and Michigan Studies and cannot be explained by roentgenographic enlargement. These differences are even the more remarkable since the Dutch data include adolescents up to 17 years of age in contrast to only 13-year-olds in both American studies. The explanation could be geographic sample differences.

Growth. The results for rate of growth per year, on the other hand, do not display remarkable discrepancies between the three studies. Taking cephalometric enlargement for the Pittsburgh and Michigan Studies into account (11–13%), the three samples appear to grow at more or less the same rate.

Number of predictors. The extension of the number of predictors from two to five apparently had no influence on the results. Age and gender estimates were equally accurate in both studies. The reason could be the low multicollinearity between the predictors. This can be verified in Tables 2, 2.1 and 2.2. Yet, this part of the operation is not fool proof and it remains the responsibility of the investigator to carefully check for the effect of increasing the number of predictors. Multiple linear regression is a sophisticated analytical tool, and while robust to departures from the underlying assumptions about its use, it still requires care in its application and statistical advice should be obtained before embarking on its use.

Conclusion

The finding in different studies and with different predictor sets that the effects of 'age', as well as the dichotomy 'gender' were for the larger part estimated correctly by the regression model strengthens our confidence in the applicability of this technique. We may assume that beta-weights will also be correctly estimated for those predictors which can not be checked, because their effect is at present unknown. It is concluded that regression analysis, applied professionally, can be a valuable analytical tool in cephalometric studies. It quantifies the association between cephalometric measures and several predictors. The creation of small subgroups is avoided because the contribution of all predictors under consideration is weighted simultaneously. Variance of major sources such as age and gender is explained, which opens the possibility to explore variance associated with more subtle predictors.

Acknowledgements

This study was supported in part by US grant HD 07403 from the National Institute of Child Health and Human Development, US grant NS 16365 from the National Institute of Neurological and Communicative Disorders and Stroke, and US grant DE 03610 from the National Institute of Dental Research.

The authors are grateful to Jack L. Paradise, MD, for providing access to the cephalograms and the data from the Children's Hospital of Pittsburg Study. Professor Dr V. Mammitzsch, University of Marburg, Germany, provided statistical counselling. D. Kathleen Colborn, BS, and Beverly S. Bernard, RN, BS, provided indispensable assistance in identifying appropriate cephalograms and in synthesising and summarising pertinent clinical data. Finally, we are grateful to Mr Richard Miller, Ms Kathleen A. O'Connor and Mr Kim Lo, Ann Arbor, Michigan, for providing considerable technical support, as well as to an anonymous referee for a constructive review.

References

Bookstein, F. L. (1982)
On the cephalometrics of skeletal change,
American Journal of Orthodontics, **82,** 177–198.

Bruin de, R. (1993)
A mathematical model applied to craniofacial growth,
Ph.D. Thesis University of Groningen, Netherlands.

Buschang, P. H., Tanguay, R., LaPalme, L. and Demirjian, A. (1990)
Mandibular growth prediction: mean growth increments versus mathematical models,
European Journal of Orthodontics, **12,** 290–296.

Cheverud, J. M. and Richtsmeier, J. T. (1986)
Finite-element scaling applied to sexual dimorphism in rhesus macaque facial growth,
Systematic Zoology, **35,** 381–299.

Cleall, J. F., Begole, E. A. and Chebib, F. S. (1979)
Craniofacial morphology: a principal component analysis,
American Journal of Orthodontics, **75,** 650–666.

Cox, D. R. and Wermuth, N. (1992)
Response models for mixed binary and quantitative variables,
Biometrika, **79,** 441–461.

Draper, N. R. and Smith, H. (1996)
Applied Regression analysis,
Wiley, New York.

Finkelstein, M., Lavelle, C. L. B. and Hassard, T. (1988)
The role of cluster analysis on traditional cephalometric dimensions,
Angle Orthodontist, **95,** 97–106.

Hallers-Tjabbes ten, C. C. (1979)
Sexual dimorphism in *Buccinum undatum* L.,
Thesis, University of Groningen, Netherlands.

Harris, J. E. (1963)
A multivariate analysis of the variability of the craniofacial complex,
Master's Thesis, University of Michigan, Ann Arbor.

Harris, J. E. (1971)
Problems in the statistical inspection of cranio-facial variables during growth and development.
In: R. E. Moyers and W. M. Krogman (Eds) *Cranio-facial Growth in Man*,
Pergamon Press, Oxford, pp. 229–231.

Howells, W. W. (1971)
Applications of multivariate analysis to cranio-facial growth.
In: R. E. Moyers and W. M. Krogman (Eds) *Cranio-facial Growth in Man*,
Pergamon Press, Oxford, pp. 209–218.

Keeling, S. D., Riolo, M. L., Martin, R. E. and Ten Have, T. R. (1989)
A multivariate approach to analyzing the relation between occlusion and craniofacial morphology,
American Journal of Orthodontics and Dentofacial Orthopedics, **95,** 297–305.

Kowalski, C. J. (1972)
A commentary on the use of multivariate statistical methods in anthropometric research,
American Journal of Physical Anthropology, **36,** 119–132.

Landauer, C. A. (1962)
A factor analysis of the facial skeleton,
Human Biology, **34,** 239–253.

Lauritzen, S. L. and Wehrmuth, N. (1989)
Graphical models for association between variables, some of which are qualitative and some quantatitive,
Annals of Statistics, **17,** 31–57.

McNamara, J. A. (1984)
A method of cephalometric evaluation,
American Journal of Orthodontics, **86,** 449–469.

Meyers, D. M. (1992)
A multivariate analysis of the relations between craniofacial structure and occlusion with DiPaolo's quadrilateral measures,
American Journal of Orthodontics and Dentofacial Orthopedics, **102,** 52–61.

Paradise, J. L. and Bluestone, C. D. (1976)
Toward rational indications for tonsil and adenoid surgery,
Hospital Practitioner, 79–87.

Riolo, M. L., Moyers, R. E., McNamara, J. A., Hunter, W. S. (1974)
An atlas of craniofacial growth,
Center for Human Growth and Development, University of Michigan, Ann Arbor.

Schneiderman, E. D., Kowalski, C. J. and Willis, S. M. (1993)
Regression imputation of missing values in longitudinal data sets,
International Journal of Bio-Medical Computing, **32,** 121–133.

Schneiderman, E. D. and Kowalski, C. J. (1994)
Analysis of longitudinal data in craniofacial research: some strategies,
Critical Reviews in Oral Biology and Medicine, **5,** 187–202.

Smillie, K. W. (1966)
An Introduction to Regression and Correlation,
Academic Press, London.

Solow, B. (1966)
The pattern of craniofacial associations. A morphological and methodological correlation and factor analysis study on young male adults,
Acta Odontologica Scandinavica, Suppl. 46.

Systat, Wilkinson, Leland (1990)
The System for Statistics,
SYSTAT Inc., Evanston, IL.

Vark van, G. N. (1985)
Multivariate analysis in physical anthropology,
Multivariate Analysis, **6,** 599–611.

Wells, D. L. (1970)
A multivariate cephalometric study of Class II, division 2 malocclusion,
Master's Thesis, University of Michigan, Ann Arbor.