MULTILEVEL ANALYSIS OF LONGITUDINAL DATA: ANALYSIS OF WORKSHOP DATA SET 1

M. MOK and R. P. McDONALD

1. INTRODUCTION

Traditional analyses of longitudinal data have typically ignored the subject effect and have assumed that the rate of change of the response variable with respect to time is constant for all subjects in the study. The models resulting from making these assumptions tend to be too rigid and interesting individual differences are neglected. Goldstein [2] provided a methodology for the use of mixed generalized linear models in handling longitudinal and repeated measures whereby growth is taken as one level of a two-level model: the within-subject model takes care of changes within individuals across occasions as a result of time and/or other explanatory factors, and the between-subject model allows for different growth rates across subjects. Raudenbush [7] further elaborated the theory and provided examples of its applications in studying school effectiveness. This paper tries to give an illustration of the use of Goldstein's mixed generalized linear model on longitudinal and repeated measures arising from a randomized group experimental design.

Goldstein [2] viewed longitudinal data as a two-level hierarchy, with subjects as the higher level (level-2) and occasions within subjects as the lower level (level-1). Goldstein suggested using a two-level polynomial growth curve model decomposing the variance into between-subjects and within-subject-between-occasion components. In this way, the response , y_{it}, of subject i on occasion t can be represented by the basic model as:

$$y_{it} = \Sigma_j \beta_{ij} x_{tj} + \Sigma_k \Omega_k z_{itk} + e_{it} \qquad (1)$$

where x_{tj}'s are some measures of time at occasion t; β_{ij}'s are usually considered as random; z_{itk}'s are explanatory variables defined at level-2; Ω_k can be random or fixed effects; $\beta_{ij} = \beta_j + \vartheta_{ij}$; e_{it} are independently

distributed with variance σ_e^2 and ϑ_{ij} has zero mean and $\text{cov}(\vartheta_{ij},\vartheta_{ij'})$ equals to $\sigma_{jj'}$.

By introducing dummy variables to the second term of equation (1), i.e. z_{itk} takes the value of one if the individual is in the kth experimental group, and zero otherwise, and by considering the corresponding coefficients Ω_k to be having fixed effects only, model (1) can be used to analyse longitudinal data from randomized block designs using hierarchical modelling.

All the analyses in this paper are carried out using the ML3 package (Version 2.2; Prosser, Rasbash, Goldstein [6]). The package provides a choice of either the iterative generalized least squares (IGLS; Goldstein, [3] or the restricted iterative generalized least squares (RIGLS; Goldstein [5]) procedures in the estimation of parameters. The analyses presented in this paper have used the default option of IGLS.

2. EXAMPLE: VITAMIN E THERAPY

The data set comes from a study which tried to investigate the effect of a Vitamin E diet supplement on the growth of guinea pigs. The body weights (gm) were taken at the end of the 1st, 3rd, 4th, 5th, 6th, and 7th weeks. At the beginning of the study, all animals were given a growth inhibiting substance and Vitamin E therapy was started at the beginning of week 5 (Hand, [1]). There were three treatment groups according to the amount of Vitamin E given. These were (1) the CONTROL group with no Vitamin E therapy; (2) the LOW group with a low dosage of Vitamin E given, and (3) the HIGH group with a high dosage of Vitamin E. There were 5 guinea pigs within each treatment group. As such, the data has a hierarchical structure (i) the animal level (level-2), and (ii) the occasion level (level-1) of repeated measurements of each animal at the 6 fixed time points.

At level-1, the 6 measurements yield a random effect across occasions within animals (within-animal effect). This within-animal effect may be a result of growth with respect to time (after the growth inhibiting drug), before- and after-treatment effects, and other random effects. At level-2, the variations can be categorised into two major types: the between animals within treatment group differences, and that between animals across treatment groups. Within each treatment

group, there were 5 guinea pigs whose weight variation may be the result of idiosyncracies of each animal including differences in initial weight, growth rate, interaction between initial weight and treatment, and other random effects. On the other hand, across different treatment groups, there is the fixed effect of treatment, and the possible effect of interaction between treatment and time such as a delayed effect of a particular treatment.

3. FITTING MODELS TO GROWTH DATA

Figure 1 shows the variations of body weight between the 15 animals. There is wide variation in the average as well as in the range of weights within each treatment group. Two mixed generalized linear models are proposed to analyse the data. For each model, the weight, y_{kit}, of observation t of the animal i in group k, is regressed on time, x_t, and treatment. The two models differ in the between-animal specifications. The first model gives a simple decomposition of variance between the two levels of animal and occasion. All the animals within the same treatment group are assumed to have the same growth rate in this model. The second model considers an additional effect on the developmental path, that of differences in growth rates for individual animals within the same treatment group.

MODEL 1: VARIANCE COMPONENTS MODEL

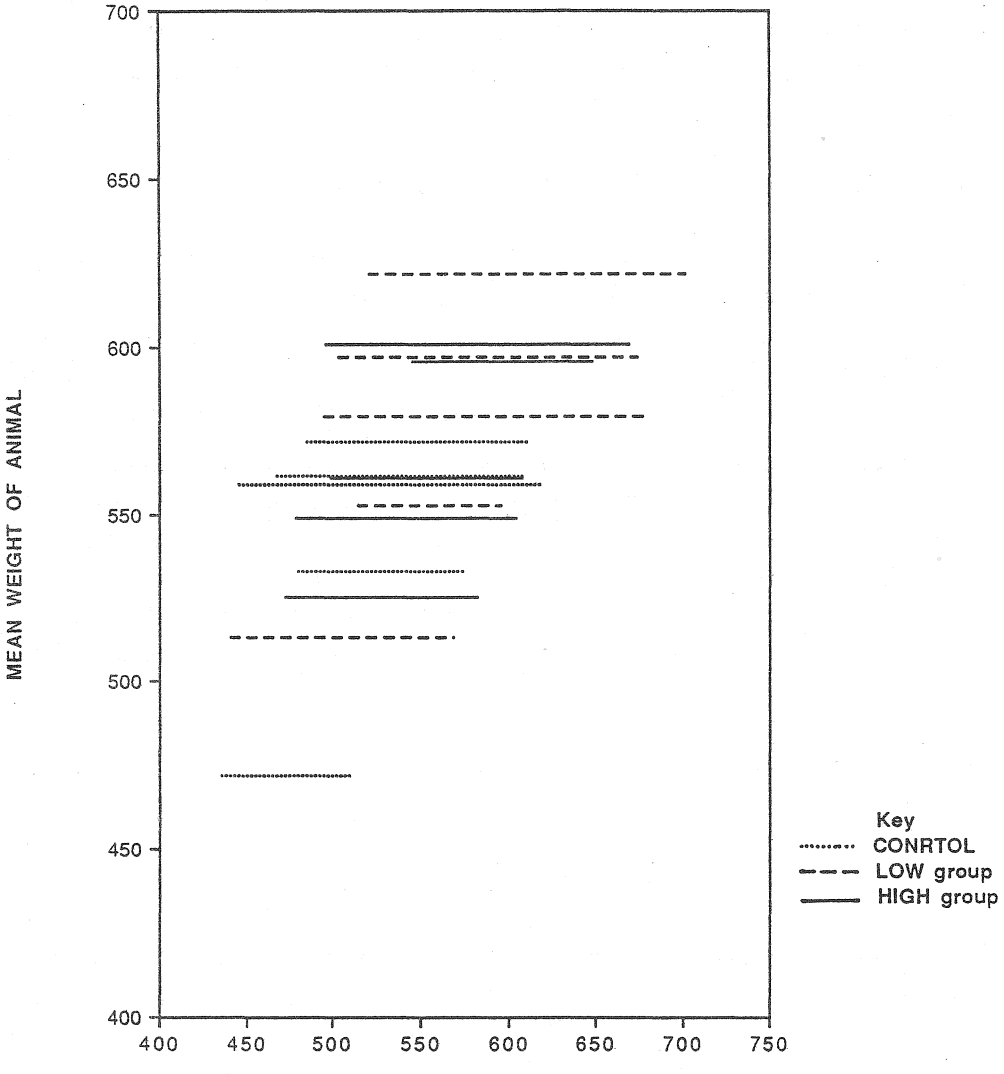
In this first model, the within-animal between occasions differences come from growth as a consequence of the mere passage of time (after the inhibiting drug at beginning of week 1), the effect of treatment at the end of week 4, and other random fluctuations from occasion to occasion. The between-animal within group effect consists of random initial weight differences. The between group effect is the fixed treatment effect operational at the beginning of week 4.

The within-animal part of the model is given by the following equation:

$$y_{kit}=ß_{0i} + ß_{1i}x_t + ß_{2ki} (x_t - 4) q_t + e_{kit} \qquad (2)$$

where y_{kit}=weight of ith animal in group k on occasion t

Fig 1.BODY WEIGHT VARIATION - 15 ANIMALS



x_t=time(week) on occasion t

$$q_t = \begin{cases} 1 \text{ if } 4 \le t \le 6 \\ 0 \text{ if } 1 \le t \le 3 \end{cases}$$

$ß_{0i}$ = initial weight of animal i

$ß_{1i}$ = growth rate of animal i

$ß_{2ki}$ = change in growth rate of animal i in group k after treatment

e_{kit} = random effect of animal i in group k on occasion t

and $e_{kit} \sim N(0, \sigma_e^2)$.

In the between-animal part of the model, the initial weight of each animal varies around the average initial weight of all animals in the study. As the animals are randomly allocated to treatment groups, before treatment there is no difference between groups and the growth rate is assumed constant across all animals. After treatment, growth is assumed to be constant within each group. The between-animal part of the model is given by the equations below:

$$ß_{0i} = \Omega_{00} + u_{oi} \qquad (3.a)$$

$$ß_{1i} = \Omega_{10} \qquad (3.b)$$

$$ß_{2ki} = \Sigma_j \ \Omega_{2j} d_j \qquad (3.c)$$

i=1,...5; j=1,...,3; k=1,...,3

where d_j=1 (if j=k) or =0 (if j≠k),

Ω_{00} = expected initial weight of all animals

Ω_{01} = expected growth rate (linear) of animal prior to treatment

Ω_{2j} = change in (linear) growth rate of animal in group j (j=1,...,3)

and,

$$\text{Var}(u_{0i}) = \sigma_0^2$$

$$\text{Cov}(e_{kit}, u_{0i}) = 0 .$$

Combining equations (2) and (3), the full variance components model can be written as:

$$y_{kit} = [\,\Omega_{00} + \Omega_{10}\, x_t + \Sigma_j\, (\Omega_{2j}\, d_j\, (x_t - 4)\, q_t\,)] + [\,u_{0i} + e_{kit}\,]\,. \qquad (4)$$

The terms within the first square bracket on the right of equation (4) are referred to as the fixed part of the model and the terms within the second square bracket are called the random part of the model. It can be easily seen that the variance matrix of the response variable y_{kit} conditional on the explanatory variables is block diagonal in form, and is given by

$$\Sigma = \sigma_e^2\, I_{90} + \sigma_0^2\, I_{15} \otimes J_6$$

where I_n is the identity matrix of order n, and J_m is an mxm matrix of ones.

RESULTS

The estimated mean growth trajectories are displayed in Table 1.

TABLE 1. RESULTS FROM A VARIANCE COMPONENTS MODEL

Parameter	Estimate	SE
(Fixed part of the combined model)		
Ω_{00}, average initial weight	459.63	11.99
Ω_{10}, average growth rate before treatment	25.84	2.68
Ω_{21}, average change in growth rate after treatment for the control group	-22.36	5.47
Ω_{22}, average change in growth rate after treatment for the low dosage group	-4.54	5.47
Ω_{23}, average change in growth rate after treatment for the high dosage group	-10.68	5.47
(Random part of the combined model)		
σ_e^2, variance of weight within animal between occasions	572.66	93.52
σ_o^2, variance of initial weight between animals	1127.4	446.7

The average initial weight is estimated to be 459.63 (SE = 11.99). The estimated average growth rate before treatment is 25.844 gm per week (SE = 2.675). It can be seen that after treatment, the changes of growth rate are all negative in sign, indicating a delayed effect of the growth inhibiting drug and that the rate of growth is slower after week 4. The decrease in growth is smallest in the low dosage group, followed by the high dosage group, and the no treatment control group slows down dramatically in growth after week 4, indicating that Vitamin E therapy might be effective in reducing the delayed growth inhibition.

It is possible to perform hypothesis testing and to compute confidence intervals for the parameters (Goldstein [4]), although it should be cautioned that with such a small sample, the asymptotic Chi-squared distribution which is used to compute the simultaneous confidence intervals might not have been attained. Tests on the hypotheses of the three slope changes being different from zero, ie

$$H_{01}: \Omega_{21}=0$$
$$H_{02}: \Omega_{22}=0$$
$$H_{03}: \Omega_{23}=0$$

give Chi-squared values of 16.72, 0.69, 3.81 respectively, each with 1 degree of freedom. Hence only the first hypothesis can be rejected at 0.1% level, which means that only the control group showed a significant decrease in growth rate after week 4; both the low dosage of Vitamin E treatment group and the high dosage group managed to counteract the effect of the growth inhibiting drug. Simultaneous 95% confidence intervals for the differences between the change in growth rate coefficients are (-28.01,-7.63) for (Ω_{21} - Ω_{22}), (-4.05, 16.33) for (Ω_{22} - Ω_{23}), and (-21.87, -1.49) for (Ω_{21} - Ω_{23}) indicating that while the two Vitamin E groups are not different from one another in treatment effect, the control group is different from each of them.

From Table 1, it can be seen that of the 1699.7 units of variation, 33.7% comes from between occasion within animal level, and 66.3% comes from the between animal level. The intra-animal correlation is 0.663, which is an indication that the multilevel approach is essential.

Figure 2 illustrates the growth path of each individual animal with the assumption that there is no variation in growth rate within groups. This is a strong assumption which might not be realistic. The next model considers the case where each animal has its own growth rate.

MODEL 2: RANDOM EFFECTS MODEL

In the random effects model, the effect of time on growth is taken to be random. The within-animal component of the model is given by the same equation (2) as in Model 1. The random effect on initial weight (Equation 3.a) and the fixed effect of treatment (Equation 3.c) are the same as before. However, an additional random effect is included in the between-animal part of the model. The between-animal part of the model is represented by the equations below:

$$ß_{0i} = \Omega_{00} + u_{0i} \qquad (5.a)$$

$$ß_{1i} = \Omega_{10} + u_{1i} \qquad (5.b)$$

$$ß_{2ki} = \Sigma_j \Omega_{2j} d_j \qquad (5.c)$$

$$i=1,...,5;\ j=1,...,3;\ k=1,...,3$$

where d_j=1 (if j=k), or =0 (if j≠k), and Ω_{00},Ω_{10},Ω_{2j} and u_{0i} are defined as in equation (3) before, and

$$\text{var}(u_{1i}) = \sigma_1^2$$

$$\text{cov}(u_{0i}, u_{1i}) = \sigma_{01}\,.$$

Coefficient u_{1i} is the random fluctuation in growth rate of animal i about the mean rate. The random effects model is given by equation (5) below:

$$y_{kit}=[\Omega_{00} + \Omega_{10}x_t + \Sigma_j(\Omega_{2j} d_j (x_t - 4) q_t)] + [u_{0i} + u_{1i}x_t + e_{kit}]\,. \qquad (5)$$

RESULTS

The results of the analysis are given in Table 2 and Figure 3. The estimates of the coefficients of the fixed part of the model are very

FIG 2. LONGITUDINAL CURVE OF WEIGHTS
VARIANCE COMPONENTS MODEL

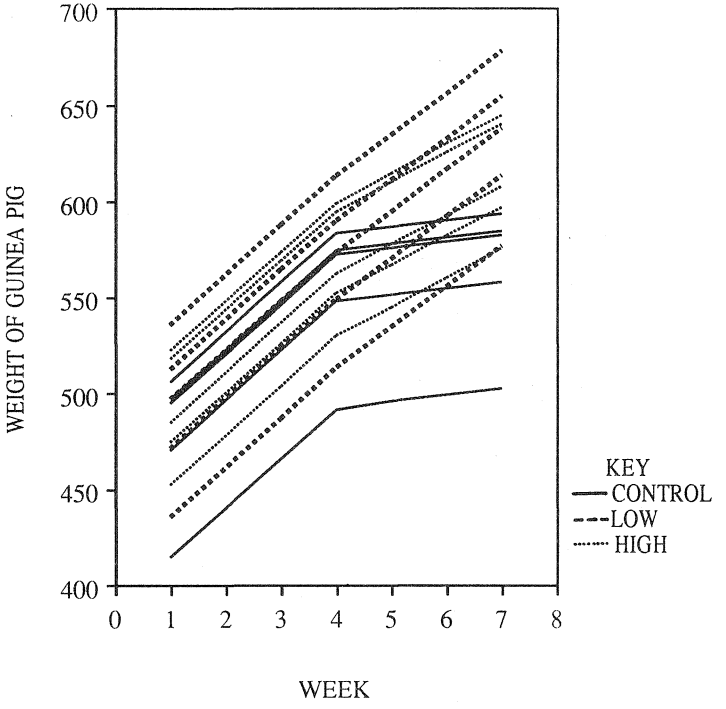
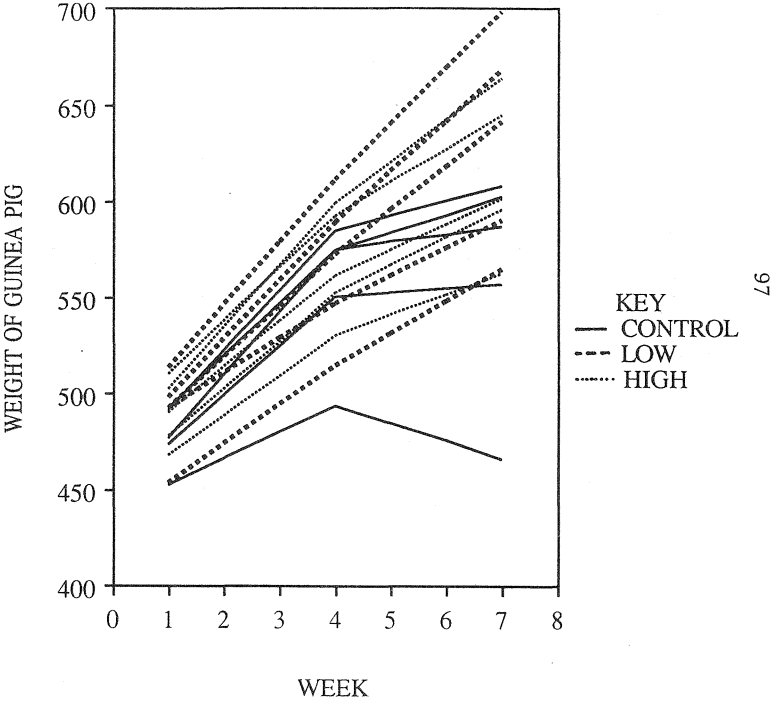


FIG 3. LONGITUDINAL CURVE OF WEIGHTS
RANDOM EFFECTS MODEL



similar to those of Model 1. The growth rate (linear) of all the groups decreased after week 4. The decrease is especially drastic in the no treatment control group (Chi-squared = 9.31, df=1, $p < 0.5\%$), resulting in almost zero growth in this group. There is no significant decrease in growth rate for the low dose group (Chi-squared = 1.05, df = 1, $p > 1\%$) and for the high dose group (Chi-squared = 4.1, df = 1, $p > 1\%$). Simultaneous 95% confidence intervals for the difference in changes in growth rate between the groups are (-31.74,-6.92) for $(\Omega_{21}-\Omega_{22})$, (-5.91, 18.91) for $(\Omega_{22} - \Omega_{23})$, and (-25.24, -0.42) for $(\Omega_{21}-\Omega_{23})$ respectively.

Intra-animal correlation can be calculated as before, and is 0.507. On the other hand, the correlation between initial weight and growth rate can be found by taking the ratio of the covariance term (9.236, SE 56.91) to the product of the square root of the respective variance estimates ($\sqrt{376.6} * \sqrt{36.77}$), which gives a value of 0.0785. The correlation is too small to justify the introduction of an interaction term between initial weight and growth rate as the potential next model.

It can be seen from Figure 3 that under the assumption of an individual growth rate for each animal, the between-treatment-group variations in the growth rates are larger than the within-treatment-group variations.

4. COMPARISON OF MODELS

It is possible to compare the variance components model (Model 1) with the random effects model (Model 2) by computing the values of the respective -2 times loglikelihood (-2*lnL) for each. The difference is distributed with a Chi-squared distribution with 2 degrees of freedom. The results show that the difference in -2*lnL is given by 17.94 (-2*lnL of Model 1 + 2*lnL of Model 2 = 865.19 - 847.255) which indicates that the random effects model is significantly better than the variance components model.

TABLE 2. RESULTS FROM A RANDOM EFFECTS MODEL

Parameter	Estimate	SE
(Fixed part of the combined model)		
Ω_{00}, average initial weight	459.63	8.56
Ω_{10}, average growth rate before treatment	25.84	2.74
Ω_{21}, average change in growth rate after treatment for the control group	-23.24	5.30
Ω_{22}, average change in growth rate after treatment for the low dosage group	-3.92	5.30
Ω_{23}, average change in growth rate after treatment for the high dosage group	-10.41	5.30
(Random part of the combined model)		
σ_e^2, variance of weight within animal between occasions	402.55	73.49
σ_o^2, variance of initial weight between animals	376.60	289.30
σ_1^2, variance of growth rate between animals	36.77	19.98
σ_{01}, covariance between initial weight and growth rate	9.24	56.91

5. CONCLUSIONS

The advantage of employing multilevel models for longitudinal and repeated measures is that the effect of characteristics specific to individual subjects can be separated from that of time. With the availability of easy-to-use packages such as ML3, such models should be used whenever appropriate to incorporate interesting features into the model.

Acknowledgement - The authors would like to expressed their sincere gratitude to Dr Chris Glasbey and Mr Gregory Robertson for discussions of an early draft of the paper.

REFERENCES

[1] Hand, D.J. (1991). Data Set 1, Longitudinal Data Analysis Workshop II, Australian National University, Canberra, July 1991.

[2] Goldstein, H. (1986a). Efficient statistical modelling of longitudinal data. Annals of Human Biology, 13, 129-141.

[3] Goldstein, H. (1986b). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73, 43-56.

[4] Goldstein, H. (1987). Multilevel models in educational and social research. London: Griffin.

[5] Goldstein, H. (1989). Restricted (unbiased) iterative generalized least squares estimation. Biometrika, 76, 622-623.

[6] Prosser, R., Rasbash, J., & Goldstein, H. (1991). ML3 Software for Three-level Analysis, Users' Guide for V.2. Institute of Education, Unviersity of London, 1991.

[7] Raudenbush, S.W. (1989). "The analysis of longitudinal, multilevel data", in Creemers, B.P.M. & Scheerens, J. (Eds) International Journal of Educational Research. Vol 13, 721-740.

Address of Authors:
Dr Mok, M., School of Education, Macquarie University, North Ryde, NSW 2109, Australia.

Professor McDonald, R. P., Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820, USA.

DISCUSSION

This paper is valuable in that it brings to our attention the components of variation that are due to subject differences. By accounting for this source of variation, the model has the potential for describing the changes with time more accurately.

However, the models here have ignored the serial correlation of the errors due to repeated measures on each subject. These correlations are a feature of data set 1 and revealed empirically by the correlogram. The variance component due to this source may dominate the modelling process.

By ignoring the serial dependence of the errors (ie by assuming {e} are independent), the authors have presented a random coefficients model that assumes equal correlation amongst the measurements (sphericity). The effect is to underestimate the variance and so overstate the significance.

The extension of the random coefficient model to incorporate the dependence of the error term is explained by DIGGLE and DONNELLY [1].

[1] DIGGLE, Peter J. and DONNELLY, John B. (1989). A selected bibliography on the analysis of repeated measurements and related areas. *Australian Journal of Statistics* **31**, 183-193.

BOB MURISON
ROSS CUNNINGHAM