Application of hierarchical linear modelling to the study of trajectories of behavioural development

MARIÁ VICTORIA HERNÁNDEZ-LLOREDA*, FERNANDO COLMENARES† & ROSARIO MARTÍNEZ ARIAS*

*Departamentos de Metodología de las Ciencias del Comportamiento and
†Psicobiología, Facultad de Psicología, Universidad Complutense de Madrid

(Received 29 January 2003; initial acceptance 6 March 2003; final acceptance 12 May 2003; MS. number: SC-1245)

Students of behavioural development often collect and analyse longitudinal data on a sample of individuals (Bateson 1981; Sackett et al. 1981; Chalmers 1987). For example, primate studies of behavioural and relationship development have typically provided data on rates of, or percentages of time engaged in, the performance of various behaviours across the individuals’ early ontogeny (e.g. Fairbanks 1996; Deputte 2000). The longitudinal records for each individual, collected on multiple occasions, are then pooled per age intervals of varying length. Commonly, such longitudinal data are analysed using either a mixed design analysis of variance (ANOVA) or a conventional (simple or multiple) regression analysis (Zar 1999; David 2002). In this commentary, we demonstrate a number of serious problems in using these methods, and describe an alternative that avoids them.

Researchers applying mixed design ANOVA to the study of behavioural development typically use age of the individual (i.e. time blocks) as the within-subject factor, and variables such as sex or species as the between-subjects factors (e.g. Maestripieri 1994; Suomi et al. 1996). Although ANOVA models can be used to analyse longitudinal data they have two major limitations. First, they require a balanced data collection design, with the same number of individuals per measurement occasion and the same interval between consecutive measurements. When these requirements are not fulfilled, which is often the case in observational studies of spontaneously occurring social interactions, a number of complications arise (e.g. Hox & Kreft 1994). Second, ANOVA models assume homogeneous (co)variances at level 1, that is, a constant variance in the parameters of longitudinal data within individuals. This assumption is often not met by developmental data, resulting in a Type I error greater than the specified α.

Animal behaviour researchers using traditional regression analysis have used two major approaches. The first treats all predictor or independent variables, such as age of infant, sex, cohort, maternal experience or dominance rank, as if they pertained to a single level (e.g. Bramblett & Coelho 1985; Altmann & Samuels 1992) and their contributions to the observed variance in the rate of behaviour during the whole study period are then assessed. The second regression modelling approach is known as the intercepts-and-slopes-as-outcomes approach (e.g. Bryk & Raudenbush 1992; Hox & Kreft 1994). In this case, two separate steps are taken. First, linear (or curvilinear) regressions are fitted to the developmental functions for each subject. Next, the parameters from these regression equations are used as dependent variables in subsequent analyses to be predicted by one or a set of independent variables (e.g. Wasser & Wasser 1995).

The application of simple or multiple, single-level linear regression to the analysis of longitudinal data is inadequate for several reasons (Bryk & Raudenbush 1992; Hox & Kreft 1994; Goldstein et al. 1998; van der Leeden 1998; Snijders & Bosker 1999). First, this approach is conceptually unsound because it ignores the hierarchical nature of longitudinal data in which observations within individuals are dependent. Second, conventional single-level methods disaggregate all higher level explanatory variables to the lowest level, followed by an ordinary least squares (OLS) regression analysis. Therefore, they do not provide methods to test for the effect of independent variables on the variation observed in the parameters of change (initial status, rate of development, and so on), which is one of the major theoretical concerns of those interested in the study of behavioural development. When OLS is applied to longitudinal data, the assumption of independence of residual error terms is violated and this leads to inefficient estimates and a Type I error rate that is much higher than the nominal α level (Hox & Kreft 1994; van der Leeden 1998).

In the intercepts-and-slopes-as-outcomes approach, separate fixed-effects regression models are first fitted within each individual, using level 1 explanatory variables as predictors. Next, the within-individual growth coefficients are, in turn, used as dependent variables to be predicted by level 2 explanatory variables (Wasser &
Wasser 1995). However, using OLS to estimate the regression coefficients in both steps is inconsistent because, in the first step, we view the regression coefficients as fixed coefficients, whereas in the second step, we view them as random variables. If the usual model assumptions are true at the lowest level, they will hardly be true at the second, and this will lead to a discrepancy between the actual error structure and the error structure assumed by the linear model. As a result, significance tests based on the usual standard errors are misleading (Hox & Kreft 1994).

The proper analysis of multilevel data requires a different class of statistical methods, which have been developed and used mainly within the field of educational statistics. Hierarchical linear models (HLMs) represent a powerful multilevel analysis technique, variously labelled multilevel models, random-coefficient models, mixed models, covariance components models (Rogosa et al. 1982; Rogosa & Willett 1985; Bryk & Raudenbush 1987, 1992; Willett 1988; Goldstein et al. 1998; van der Leeden 1998; Snijders & Bosker 1999; Raudenbush 2001). When applied to longitudinal data, these allow the estimation of each individual’s development trajectory, a group’s average development trajectory and individual variation around the average trajectory; they also allow the analysis of correlates and predictors of individual differences in developmental pathways.

Our main aim in this paper is to draw the attention of animal behaviour researchers to hierarchical or multilevel analysis methods and to show when and why they are more useful and appropriate for studying behavioural change and developmental trajectories than the mixed design ANOVA or the traditional regression analysis. We provide a brief description of the hierarchical linear modelling framework in the context of the study of behavioural development and, by way of example, we apply it to the analysis of the time course of development of infant carrying by the mother in a colony of hamadryas baboons, *Papio hamadryas*, during the infant’s first year of life. Finally, we apply the other two conventional regression analysis methods described above to the same data set to compare them with the multilevel methodology.

Hierarchical Linear Modelling of Longitudinal Data

In hierarchical linear modelling, any number of levels can be represented, but the essential statistical features are found in the basic two-level models. These are briefly outlined below.

The application of a two-level growth modelling approach to the study of behavioural development involves the specification of a pair of linked statistical models: one for individual development (i.e. level 1, within-subject or repeated-observations model) and one for interindividual differences in development (i.e. level 2, between-subjects or individual-level model). At the first stage (level 1), each individual’s observed development trajectory is conceived of as a function of an individual growth trajectory plus random error. This trajectory is determined by a set of individual parameters. At the second stage (level 2), we assume that these individual development parameters vary randomly or as a function of differences between individuals in a number of variables (e.g. physiological, organismic, psychological, social, experiential and ecological factors). This two-stage conceptualization of development implies the need for a model in which the parameters in the first stage become the outcome measures in the second (Bryk & Raudenbush 1987, 1992; Willett 1988; Snijders & Bosker 1999). This hierarchical dependence between parameters at level 1 and level 2 is precisely what makes the model hierarchical or multilevel (Raudenbush 2001).

The within-subject model

Individual development may be modelled as a polynomial function of time. When the investigator knows little of the processes that underlie development, this approach is recommended (Willett 1988). In general, we assume that $Y_{it}$, that is, the measurement on the dependent or outcome variable for individual $i$ on occasion $t$, is a function of a systematic growth trajectory or development curve plus random error. Thus, the basic within-subject (level 1) model could be written as:

$$Y_{it} = \beta_{0i} + \beta_{1i}T_{it} + \beta_{2i}T_{it}^2 + \ldots + \beta_{pi}T_{it}^p + e_{it} \tag{1}$$

The $\beta$s are the coefficients of a polynomial function of degree $p$ and the $e_{it}$ are random error terms. The level 1 explanatory or predictor variable $T_{it}$ could represent, for instance, the age of individual $i$ at time $t$. The powers $\beta_{gi}T_{it}^g$ ($g=1, \ldots, p$) represent transformations of this variable, specifying a quadratic, cubic or higher order polynomial development function (Bryk & Raudenbush 1992; Bryk et al. 1996; van der Leeden 1998). When the data available have been collected over a long period it may be appropriate to select a curvilinear growth model to represent each subject’s individual development trajectory. In a second-order or quadratic regression (i.e. second-degree polynomial), for example, we would have three individual development parameters: the initial status (i.e. intercept), $\beta_{0i}$, and the coefficients of the linear and quadratic terms, $\beta_{1i}$ and $\beta_{2i}$, which can be taken to represent development rate and acceleration, respectively (Willett 1988; Bryk & Raudenbush 1992).

The between-subjects model

A key feature of equation (1) is the assumption that the individual growth parameters (i.e. the $\beta$s) vary across subjects. At this stage, we have to formulate a between-subjects model to represent this observed heterogeneity in individual development trajectories. As already mentioned, the $\beta$ parameters in the within-subject model become the outcome variables in the level 2 model. A simple between-subjects model, with simple random variation in the individual development parameters across the group, could be:

$$\begin{align*}
\beta_{0i} &= \gamma_{00} + u_{0i} \\
\beta_{1i} &= \gamma_{10} + u_{1i} \\
\vdots \\
\beta_{pi} &= \gamma_{pi0} + u_{pi}
\end{align*} \tag{2}$$
In equation (2), the parameters, $\gamma_{00}, \gamma_{10}, \ldots, \gamma_{p0}$ of the between-subjects model are known as fixed effects, and represent the average development curve over all individuals. The $u_{0i}, u_{1i}, \ldots, u_{pi}$ are random error components denoting the departures of $\beta_{0i}, \beta_{1i}, \ldots, \beta_{pi}$ from these $\gamma_{00}, \gamma_{10}, \ldots, \gamma_{p0}$ coefficients for each individual $i$. This level 2 model of simple random variation is called an unconditional model since no level 2 predictors for any of the $\beta$ parameters have been introduced (Bryk & Raudenbush 1992; van der Leeden 1998). More elaborate between-subject models can be developed if one or several level 2 predictor variables are added to the model to try to explain part of the variability of the level 1 $\beta$ coefficients. In that way, it is possible to analyse the variability of the development curve coefficients as a function of the differences between individuals on some characteristic that remains fixed across occasions (e.g. the subject’s sex). For example, we may try to account for the variation in individual development trajectories using a level 2 variable $Z$, and then the new between-subjects model could be written as:

$$
\beta_{0i} = \gamma_{00} + \gamma_{01}Z_i + u_{0i},
$$

$$
\beta_{1i} = \gamma_{10} + \gamma_{11}Z_i + u_{1i},
$$

\vdots

$$
\beta_{pi} = \gamma_{p0} + \gamma_{p1}Z_i + u_{pi}
$$

In equation (3), the variation in individual development parameters (i.e. $\beta_{0i}, \beta_{1i}, \ldots, \beta_{pi}$) is taken to be a function of differences between individuals in the level 2 predictor variable $Z$, and the coefficients $\gamma_{01}, \gamma_{11}, \ldots, \gamma_{p1}$ stand for the average effect of variable $Z$ on the individual development parameters (van der Leeden 1998). Any number of $Z$ variables can be included as predictors in the model, the criterion for inclusion being the magnitude of the estimated effect of a given predictor and the associated $t$ ratio (Bryk & Raudenbush 1992). When level 2 predictor variables are incorporated into the model this becomes a conditional model (Bryk & Raudenbush 1992).

Although separate models for each level may be formulated, they are in fact connected statistically (van der Leeden 1998). Estimation of each of these depends on the others (Bryk & Raudenbush 1992; Bryk et al. 1996). Available programs compute both ordinary least squares (OLS) and empirical Bayes estimates (EB) of the estimated effect of a given predictor and the associated $t$ ratio (Bryk et al. 1996). To run all the analyses we used the software packages HLM for Windows, version 4.01 (Bryk et al. 1996), MlwiN version 1.00 (Goldstein et al. 1998) and SPSS version 11.0.

**Results and discussion**

**Multilevel approach.** Visual inspection of the data on the percentage of time that the infants were transported ventrally by their mothers during their first year of life suggested fitting curvilinear growth models to represent true individual development. The final model fitted to the VT data was a quadratic growth model. Adding a third (or cubic) term to the model did not improve the fit as it was nonsignificant ($\gamma_{30} = 0.000028, P = 0.134$). The other three estimated mean growth parameters did reach statistical significance, that is, the initial status ($\gamma_{10}$), the development rate ($\gamma_{10}$) and the acceleration rate ($\gamma_{20}$), thus indicating that all three parameters are necessary for mathematically describing the average development trajectory of VT (Table 1). The subjects differed significantly from one another in the time that they were carried ventrally by their mothers during the first 2-week period of age ($U_i$: initial status) and in the rate of change of this measure afterwards ($U_i$: development rate; Table 1). At age interval 1 (first 2 weeks of life), the infants spent on average about 22% of their time being transported ventrally; from then onward ventral riding steadily declined with time, and by month 9 it was essentially over (Fig. 1).

Of the three potential predictors at level 2, only mother’s age and mother’s reproductive experience fulfilled the criterion for inclusion in the model (i.e.
t ratio > 1; Table 2). Owing to limits imposed by the number of level 2 units and the multicollinearity problems that can arise, Bryk & Raudenbush (1992) recommended that, to determine which predictors will enter the final model, separate submodels are first fitted for each candidate predictor. These submodels are then combined into an overall model if the level 2 predictors examined (i.e. mother’s age and reproductive experience) do account for the variance observed in the dependent variable. In this study, however, we found that only the predictor mother’s reproductive experience had a statistically significant effect and, therefore, was the only level 2 predictor included in the final model. The mother’s reproductive experience had a statistically significant effect on both the amount of time that the infants were transported ventrally at age interval 1 and on the rate of development of this behaviour (Table 3). Thus, high-experience mothers spent significantly less time transporting their infants ventrally than low-experience mothers did and, in addition, the former’s rate of change was smoother than that of low-experience mothers (Fig. 1c). The mother’s reproductive experience accounted for 24% of the variance in the initial status and 50% of the variance in the rate of change, which showed a decreasing trend.

The quadratic growth model fitted to the data on VT accounted for 71% of the level 1 variance ($R_1^2=0.71$) and for 67% of the level 2 variance ($R_2^2=0.67$). ($R_1^2$ is the proportional reduction in mean squared prediction error, for the prediction of a given individual’s score at any given point in time. $R_2^2$ is the proportional reduction in mean squared prediction error, for the prediction of an individual’s mean score across time.)

The four model assumptions were fulfilled by the data.

ANOVA approach. We could not apply this analysis method to our VT data because when we tried to meet one of the ANOVA requirements, that is, a balanced data design, our sample size was too small. In effect, we had only six subjects that contributed data in all the 25 2-week intervals. In general, the multilevel methodology is much more flexible than are approaches based on ANOVA because both the number of observations per

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status, $\beta_0$</td>
<td>Mean initial status, $\gamma_{00}$</td>
<td>0.5322</td>
<td>0.0266</td>
<td>19.94</td>
</tr>
<tr>
<td>Growth rate, $\beta_1$</td>
<td>Mean growth rate, $\gamma_{10}$</td>
<td>-0.0437</td>
<td>0.0027</td>
<td>-15.76</td>
</tr>
<tr>
<td>Acceleration, $\beta_2$</td>
<td>Mean acceleration, $\gamma_{20}$</td>
<td>0.0009</td>
<td>0.00009</td>
<td>9.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Variance component</th>
<th>$\chi^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status (Intercept: $U_0$)</td>
<td>0.01039</td>
<td>138.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Growth rate (Age interval: $U_1$)</td>
<td>0.00002</td>
<td>69.05</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level 1 residual ($\epsilon_u$)</td>
<td>0.00834</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Status (%) of ventral transport (VT) as a function of infant age over the first year of life. (a) Estimated individual growth trajectories for each of the 23 dyads. (b) Estimated average development trajectory for the whole sample ($N=23$ mother–infant dyads). (c) Estimated average development trajectories for mothers varying in reproductive experience. The relation between infant age and the amount of VT was plotted with reproductive experience held constant at two values, one infant (lower curve) and 10 infants (upper curve).
individual and the spacing of these observations in time can vary without posing any special problem (Bryk & Raudenbush 1987, 1992; Goldstein et al. 1998; van der Leeden 1998; Snijders & Bosker 1999). Hierarchical linear models are also more flexible than ANOVA methods in that the former allow for the modelling of a variety of within-subject covariance structures (Bryk & Raudenbush 1992; van der Leeden 1998).

**Single-level approach.** Conventional, single-level, multiple stepwise regression analysis was used to investigate the relative contribution of age and sex of infant, and mother’s age and reproductive experience, to the observed variation in the percentage of time spent in VT. The best fit model to account for the variation in VT was defined by a second-degree polynomial or quadratic regression. The total variance accounted for by the model was 70% ($R^2 = 0.70$). The percentage of time spent in VT was correlated with infant age ($r^2 = -0.62, P < 0.001$), infant sex ($r^2 = -0.01, P = 0.02$) and mother’s reproductive experience ($r^2 = -0.02, P < 0.001$). (Maternal age did not enter in the equation because it was highly correlated with the mother’s reproductive experience; $r_{21} = 0.76, P < 0.001$).

One of the central assumptions of standard regression analysis is that the random errors are independent. This assumption is hard to make when dealing with longitudinal data and was not fulfilled by our own data. As already pointed out, single-level, multiple regression analysis has various shortcomings. Violation of the assumption of independent errors, as with our data, is known to lead to inefficient estimates and a Type I error rate that is higher than the nominal alpha. This could be a major reason why the two models yielded different results. Thus, the single-level model fitted to the VT data included three statistically significant predictors, infant age, infant sex and mother’s reproductive experience, whereas in the fitted multilevel model, the effect of infant sex was nonsignificant. In addition, the single-level model accounted for the effect of the predictors on the variation in level of VT but not, as did the multilevel model, on the variation in the development trajectory of VT.

**Intercepts-and-slopes-as-outcomes approach.** Once the developmental functions for each subject had been calculated, based on ordinary least square regression analysis, the coefficients from these regression equations were used as dependent variables (i.e. initial status of the behaviour at birth, and rate and acceleration of development across time) and its relationship with the independent variables was assessed with multiple regression analysis. As regards the variation observed in the initial status of VT, the regression equation included infant sex and mother’s reproductive experience as predictor variables ($t = -2.56, P = 0.02$ and $t = -2.81, P = 0.01$).

<table>
<thead>
<tr>
<th>Potential level 2 predictors</th>
<th>Level 1 coefficient</th>
<th>Intercept</th>
<th>Growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>t</td>
</tr>
<tr>
<td>Sex of infant</td>
<td>-0.330</td>
<td>0.390</td>
<td>-0.38</td>
</tr>
<tr>
<td>Mother’s age</td>
<td>-0.001</td>
<td>0.000</td>
<td>-2.14</td>
</tr>
<tr>
<td>Mother’s reproductive exp.</td>
<td>-0.014</td>
<td>0.005</td>
<td>-2.69</td>
</tr>
</tbody>
</table>

**Table 2. Exploratory analysis of t ratios to determine the level 2 predictor variables that are candidates to be included in the final model.**

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>For initial status, $\beta_0$</td>
<td>Mean initial status, $\gamma_{00}$</td>
<td>0.5938</td>
<td>0.03450</td>
<td>17.20</td>
</tr>
<tr>
<td>Effect of MRE $\gamma_{01}$</td>
<td>-0.0161</td>
<td>0.00646</td>
<td>-2.49</td>
<td>0.02</td>
</tr>
</tbody>
</table>

| For growth rate, $\beta_1$ | Mean growth rate, $g_{10}$ | -0.0461 | 0.00297 | -15.53 | <0.001 |
| Effect of MRE $\gamma_{11}$ | 0.0006 | 0.00030 | 2.05 | 0.05 |

**Table 3. Quadratic (conditional) model of development of ventral transport (effect of mother’s reproductive experience, MRE).**

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Variance component</th>
<th>$\chi^2_{21}$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status (Intercept: $U_0$)</td>
<td>0.00789</td>
<td>101.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Growth rate (Age interval: $U_1$)</td>
<td>0.00001</td>
<td>54.37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level 1 residuals ($\epsilon_0$)</td>
<td>0.00835</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
respectively). However, none of the variables accounted for the variation observed either in the rate (i.e. slope) or in the acceleration (i.e. quadratic term) of the development of VT.

As in the single-level regression analysis, in the intercepts-and-slopes-as outcome model the assumption of independent errors was again violated. This may account for the different results from those obtained by hierarchical modelling. Of the three parameters describing the individuals’ development trajectories, only the initial status of VT was accounted for by some of the predictors examined. In fact, the conventional regression model included two variables, infant sex and mother’s reproductive experience, as predictors of the variation in the status of the individuals’ VT in the first 2 weeks of life. This is in contrast with the findings obtained by hierarchical linear modelling, where variation both in initial status and in rate of development was explained by differences in the mothers’ reproductive experience.

Conclusion

Hierarchical linear regression modelling already has a fairly long tradition in the field of education and social research (Willett 1988; Bryk & Raudenbush 1992; Raudenbush 2001). We believe that this statistical methodology should be equally useful in the hands of animal behaviour researchers interested in the study of behavioural and relationship change. Multilevel modelling represents a powerful statistical tool to tackle effectively many of the conceptually and theoretically relevant questions frequently posed by students of behavioural and social development, including the mathematical description of an individual’s development trajectory, the comparison of different individuals’ development trajectories, and the investigation and adequate testing of hypotheses about possible underlying processes. Hierarchical linear regression models take fully into account the multilevel structure of longitudinal data, are flexible regarding both the handling of unbalanced designs and the modelling of variable covariance structures, and enable researchers to integrate in a single model the developmental study of normative processes and of individual differences.

We are very grateful to María José Hernández Lloreda, M. A. García Pérez and Vicente Nuñez-Antón for advice, and three anonymous referees for their detailed and valuable comments that have improved the manuscript. Portions of this research were partly supported by a postdoctoral fellowship from the MEC (Spain), 1988–1991, and project grant PB98-0773 from the MCyT (Spain) to F.C.

References


