Multilevel multivariate modelling of childhood growth, numbers of growth measurements and adult characteristics

Harvey Goldstein and Daphne Kounali

University of Bristol, UK

Summary. A general latent normal model for multilevel data with mixtures of response types is extended in the case of ordered responses to deal with variates having a large number of categories and including count data. An example is analysed by using repeated measures data on child growth and adult measures of body mass index and glucose. Applications are described that are concerned with the flexible prediction of adult measurements from collections of growth measurements and for studying the relationship between the number of measurement occasions and growth trajectories.

Keywords: Latent normal model; Mixed response types; Multilevel; Multivariate multilevel model; Ordered categories; Ordered probit model; Poisson latent normal transformation; Prediction

1. Introduction

In studies of childrens' growth a common objective is to develop procedures that can be used to predict adult conditions from repeated measurements taken during childhood. Early attempts to do this for adult variables such as stature essentially consider linear prediction models for the response regressed on a set of growth measures taken at a predefined set of ages (usually at whole years of age—see, for example, Tanner *et al.* (2000)). Such a system, however, lacks flexibility since in practice individuals will be measured at a set of ages that do not typically correspond to the predetermined set. A more flexible approach is to model the growth measures as functions of age, jointly with the adult measurements, by using suitable longitudinal data. The resulting model parameters can then be used to provide a suitable prediction model that will apply, in principle, to any set of growth measurements taken on an individual. In the present paper we consider such a model applied to the prediction of adult glucose levels and body mass index BMI (measured in kilograms per square metre), using measures of weight taken in children under the age of 10 years. These predictions, together with suitable prediction intervals, can then be used as a screening device for the early detection of later problems.

The basic design, therefore, is based on a two-level repeated measures model as described, for example, in Goldstein (2003), chapter 5. Level 1 is defined by the growth measurement occasions and these are clustered within each individual, which is the level 2 unit. The idea is that the growth measurement parameters describing average growth, velocity of growth

Address for correspondence: Harvey Goldstein, Graduate School of Education, University of Bristol, Bristol, BS1 8JA, UK.
E-mail: h.goldstein@bristol.ac.uk

etc. vary at the individual level and covary among themselves and with the adult outcomes. If we assume multivariate normality, the parameters of the fitted model, notably the elements of the individual level covariance matrix, then provide sufficient statistics for the linear prediction of any of the adult measurements from any combination of growth measurements.

One of the problems in repeated measures growth studies is that attrition may be non-random, leading to parameter estimation bias. A further aim of the present paper is to study whether this might be so. A standard procedure for dealing with informative dropout is via mixture models (Hogan and Laird, 1997) where the response vector, i.e. the set of growth measures, and the dropout distribution are jointly modelled. In the present paper we study the issue of informative dropout by modelling the actual number of occasions on which each individual was measured, jointly with the growth parameters and with adult measurements, providing estimates of the relevant correlations.

In the next section we describe a general procedure for these joint models and show how it can be specialized to handle our data.

2. Modelling mixed response types at two levels

Goldstein *et al.* (2009) present a general model (the GCKL model) for multivariate responses that can occur at several levels of a data hierarchy and where the responses can be continuous or discrete. The basic methodology is described in a freely downloadable document from the Centre for Multilevel Modelling Web site (Goldstein *et al.*, 2007): `http://www.cmm.bristol.ac.uk`.

This work includes, as special cases, existing literature on the joint modelling of discrete and continuous responses, and this literature is referenced there. The GCKL model extends previous work by allowing for a multilevel structure for the data and can also jointly model ordered, unordered and continuous data. The ability to model ordered responses within the GCKL framework provides a very general procedure for any kind of discrete data where we can treat the categories as an ordering, including count data. In the present paper we explore the case where we have a mixture of normal and count data at two levels of a data hierarchy. If we treat count data as an ordered classification Goldstein *et al.* (2009) provide a procedure to obtain parameter estimates, but this may be inefficient and even numerically unstable, when the number of categories is very large, since it then involves the estimation of a large number of 'threshold' parameters. We therefore propose two extensions to the GCKL model: one involves using smoothing functions for the large number of threshold parameters that are involved in the algorithms that were proposed by Goldstein *et al.* (2009) and the other is based on a latent normal characterization for the Poisson distribution that can model count data that can be assumed to follow such a distribution. Since the growth measure in our data is restricted to weight, it may in practice have limited utility as a sensitive prediction model, but the methodology that we propose will have general applicability.

We now briefly review the two-level GCKL model as it applies to normal and ordered responses. Full details are given in Goldstein *et al.* (2009). We begin by describing the multivariate normal model and its estimation and in the following section we show how ordered responses can be dealt with by using a transformation to underlying normality.

2.1. The GCKL latent normal model
Let $j = 1, \ldots, J$ index level 2 units and $i = 1, \ldots, I_j$ index level 1 units, nested within the level 2 units. The underlying multivariate normal model is

$$\left. \begin{array}{c} y_{ij}^{(1)} = X_{1ij}\beta^{(1)} + Z_{1ij}u_j^{(1)} + e_{ij}^{(1)}, \\ y_j^{(2)} = X_{2j}\beta^{(2)} + Z_{2j}u_j^{(2)}, \\ e_{ij}^{(1)} \sim \text{MVN}(0, \Omega_1), \quad u_j = (u_j^{(1)}, u_j^{(2)})^{\mathrm{T}}, \quad u_j \sim \text{MVN}(0, \Omega_2). \end{array} \right\} \tag{1}$$

The superscripts denote the level at which a variable is measured or defined. Here $y_{ij}^{(1)}$ is a p_1 row vector containing the normal responses that are defined at level 1 for level 1 unit (observation) i nested in level 2 unit j. Without loss of generality, we assume the same set of predictors for each of these level 1 responses. Let X_{1ij} be an f_1 row vector that contains observations on these predictor variables for observation i nested in higher level unit j and $\beta^{(1)}$ is an $f_1 \times p_1$ matrix containing their fixed coefficients. The columns of $\beta^{(1)}$ correspond to coefficients for each level 1 response variable. Similarly Z_{1ij} is a q_1 row vector that contains for observation i nested in higher level unit j values of q_1 predictor variables related to the q_1 random effects for the level 1 responses. Quite often they may be a subset of the X_{1ij}. The $u_j^{(1)}$ is a $q_1 \times p_1$ matrix containing the random effects at level 2 for the level 1 responses, with each column relating to a particular response variable. The random effects may be correlated with each other both within and across level 1 responses. Each response has an associated level 1 residual and these are contained in the p_1 row vector $e_{ij}^{(1)}$. Their covariance matrix is Ω_1.

Correspondingly, in the second part of the expression $y_j^{(2)}$ is a p_2 row vector containing the remaining responses that are defined at level 2 and X_{2j} is an f_2 row vector that contains predictor variables for higher level unit j, again assumed to be the same for each level 2 response variable. The fixed coefficients of these variables are contained in the $f_2 \times p_2$ matrix $\beta^{(2)}$. The Z_{2j} is a q_2 row vector related to the level 2 random effects for the level 2 responses. The $u_j^{(2)}$ is a $q_2 \times p_2$ matrix of correlated level 2 residuals which are also correlated with the level 2 residuals for the level 1 response. The full covariance matrix of all the level 2 residuals is Ω_2.

In the examples of this paper we assume that $q_2 = 1$. We shall also assume that the level 1 responses only have simple intercept random effects.

Estimates for the parameters of this model are obtained by using Markov chain Monte Carlo sampling and details are given in Goldstein *et al.* (2009). Assuming starting values and a burn-in period, given current values, the sampling steps are outlined as follows.

Step 1: sample a new set of fixed coefficients at level 1.
Step 2: sample a new set of fixed coefficients at level 2.
Step 3: calculate a new set of level 2 residuals for level 2 responses by subtracting the fixed predictor $X_{2j}\beta^{(2)}$ from the associated normal responses.
Step 4: sample a new set of level 2 residuals for level 1 responses.
Step 5: calculate a new set of level 1 residuals by subtracting the prediction using the fixed effects plus the level 2 random effects, $x_{1ij}\beta^{(1)} + Z_{1ij}u_j^{(1)}$, from the associated normal responses.
Step 6: sample a new level 1 covariance matrix.
Step 7: sample a new level 2 covariance matrix.

Missing responses are treated by imputing the corresponding random effect conditional on current values for all the fixed predictors and correlated random effects (Goldstein *et al.*, 2009). In the application that we consider there are no missing responses for the level 2 (individual) unit measurements. We also note that, in the framework of a repeated measures model, there is no requirement to have the same number of level 1 (measurement occasion) units for each level 2 unit (Goldstein (2003), chapter 5).

There are particular issues about the choice of default diffuse prior distributions for the covariance matrices. For example, Goldstein *et al.* (2009) describe the choice of independent uniform priors for the elements of this matrix, and Browne (2006) suggested that, where max-

imum likelihood estimates for parameters are readily available from a preliminary fit, it may be reasonable to use an inverse Wishart prior with covariance parameter equal to the corresponding maximum likelihood (or quasi-likelihood) estimate and with degrees of freedom equal to the order of the covariance matrix. We could also use an 'adaptive prior' whereby a set of iterations during a secondary burn-in period is averaged and this is used as the prior for a further burn-in period and the final set of iterations. In the examples of the present paper, because we shall be dealing with constrained covariance matrices, we shall need to update the elements of these matrices one by one and this leads us to use a set of independent uniform priors for these elements.

The GCKL model also allows for non-normal responses by formulating a latent normal model where an extra step is inserted to sample an underlying (latent) normal variable from the observed response. Goldstein *et al.* (2009) show how this can be done for ordered and unordered categorical data. They additionally allow for non-normal continuous data by using a Box–Cox transformation. When such a latent normal variable is sampled, in addition to conditioning on the current values of the parameters and the observations it is also necessary to condition on the correlated responses and this ensures that the latent variables and observed normal variables have a joint multivariate normal distribution at both level 1 and level 2. In the next two sections we outline the procedure for general ordered data and then describe the extension to Poisson data. In all these cases, for ease of exposition, we consider the single-level case. The extension to two levels follows the description that was given above.

3. Sampling a latent normal variable for an ordered response

3.1. A general procedure

Suppose that we have an ordered p-category response indexed by numbers $g = 1, \ldots, p$. For simplicity we shall use a single-level index i and denote by π_{gi} the probability that observation i occurs in category g. The cumulative distribution is defined as

$$\gamma_{hi} = \sum_{g=1}^{h} \pi_{gi}, \qquad h = 1, 2, \ldots, p-1, \text{ with } \gamma_{pi} = 1.$$

We consider the probit link model for these as

$$\gamma_{hi} = \int_{-\infty}^{\alpha_h - (X\beta)_i} \varphi(t) \, \mathrm{d}t, \tag{2}$$

where $\varphi(t)$ is the density function of the standard normal distribution and $(X\beta)_i$ is the value for i of a linear predictor expressed in matrix terms. The α_h are known as threshold parameters. In general the linear predictor $(X\beta)_i$ will contain higher level random effects and also the terms that arise from conditioning on all correlated random effects from the remaining responses. Since the joint distribution is multivariate normal, as pointed out above, this conditioning is linear and is readily derived from the current value of the covariance matrix. Details are given in Goldstein *et al.* (2009).

We also note that in equation (2) the linear predictor remains constant over the categories that are defined by the thresholds. Interactions between the threshold parameters and the predictor variables are possible and can be included in straightforward ways.

We can also motivate the latent normal model (2) as follows. If y_i is the observed category then $y_i \leqslant h$ if and only if a latent variable $y_i^* \leqslant 0$ where $y_i^* = (X\beta)_i - \alpha_h + e_i$, $e_i \sim N(0, 1)$. Thus, we have

$$\mathrm{pr}(y_i \leqslant h) = \mathrm{pr}(y_i^* \leqslant 0) = \mathrm{pr}\{e_i \leqslant \alpha_h - (X\beta)_i\} = \int_{-\infty}^{\alpha_h - (X\beta)_i} \varphi(t)\,\mathrm{d}t.$$

If an intercept term is incorporated in the fixed part predictor then there must be a linear constraint on the α_h and it is convenient to take $\alpha_1 = 0$ for this.

We can convert the categorical response to the latent normal model by sampling as follows.

(a) For a category 1 response we sample randomly from the standard normal distribution restricted to $[-\infty, -(X\beta)_i]$.
(b) For a category p response we sample from the standard normal distribution restricted to $[\alpha_{p-1} - (X\beta)_i, \infty]$.
(c) For every other category h we sample from the standard normal distribution restricted to $[\alpha_{h-1} - (X\beta)_i, \alpha_h - (X\beta)_i]$.

Sampling for the threshold parameters α_h is most efficiently carried out by using a Metropolis step (Goldstein *et al.*, 2009).

3.2. Smoothing the threshold parameters

As mentioned in Section 1, for many kinds of count data, if we treat every observed count value as a separate category, we may have a very large number of threshold parameters to estimate, and furthermore some of the categories may be small and these can cause computational problems. One solution would be to merge adjacent categories where there are small numbers. This, however, suffers from several drawbacks.

First, if there are categories with few observations the associated threshold parameters will not be well estimated and the computational time will be increased. Secondly, although the merging of adjacent categories will not change the parameters being estimated in an additive linear model, this is not so if there are interactions between the threshold parameters and the linear model coefficients. If such interactions exist and are ignored then in general we shall have different values for these coefficients depending on which categories are merged. Likewise, if an interactive model is fitted, and categories involved in these interactions are merged, then different values generally will be estimated. Thirdly, if adjacent categories are merged and this is felt to be justified in terms of the underlying model, it will not allow us to make inferences about the proportions that are associated with the original, unmerged, categories. This restriction does not apply to the smoothing approach that is now described.

In this approach we seek a functional form to describe the threshold values based on a few parameters and the following describes one such approach.

In equation (2) write

$$\alpha_h = f(h), \qquad h = 1, \ldots, p-1, \tag{3}$$

where f can be any suitable function. We shall illustrate by using a simple polynomial as a regression function of the category sequence number h typically centred at the mean value. Other choices are possible, including fractional polynomials and regression splines, and we could also use, for example, a set of starting values computed for the thresholds that are derived directly from assigning normal equivalent deviates to the cumulative category probabilities, but we have not investigated these possibilities. A further possibility that is currently being explored (Goldstein, 2008) is a function that consists of a sum of exponential terms that automatically satisfies the monotonicity requirement (see below). We have, omitting subscripts,

$$\alpha_h = \delta_0 + \delta_1 h + \delta_2 h^2 + e. \tag{4}$$

We sample each δ in turn by using a Metropolis step with a normal proposal. Using equation (4), with the α_h so determined, we then can compute the set $\{\alpha_h\}$ and thus the likelihoods for acceptance or rejection as in the standard algorithm for ordered data (Goldstein *et al.*, 2009). The function f must be strictly increasing over the range of h and this is easily verified at each step: if the order constraints for the set of thresholds are not obeyed we do not update the value. We then sample from the latent normal distribution as in the ordered case as detailed above.

To obtain starting values we first obtain starting values for each separate α as described in Goldstein *et al.* (2009), and then we fit the regression

$$\alpha_h = \sum_{t=0}^{q} \delta_t h^t + e_h \tag{5}$$

to determine the order of polynomial to be fitted and starting values for coefficients. The estimated standard errors for the coefficients can be used for the normal proposal distributions.

Finally, it is possible to have a partially smoothed model where the smoothing takes place only over a contiguous subset of the threshold parameters, e.g. those where numbers are small.

3.3. Count data

An alternative to the previous section's treatment of count data is to assume a particular parametric distribution for the counts and we shall consider the common one, namely that the counts follow a Poisson distribution. If a Poisson distribution can be supported by the data then it will be efficient to model the data explicitly assuming such a distribution, since we shall only need to estimate a single parameter. As with general ordered data we can then incorporate this distribution within the GCKL framework that underpins the procedures which are developed in this paper. To do this we need to find a Markov chain Monte Carlo step to sample a latent normal variable and we now show how this can be done.

Consider the Poisson density function

$$f(h; \theta) = \exp(-\theta)\theta^h / h!, \qquad h = 0, \ldots, p-1, \tag{6}$$

for the first p categories that are observed.

As in equation (2) we write the cumulative distribution

$$F(h; \theta) = \sum_{g=0}^{h} f(g; \theta).$$

We choose a reference value $h = h_0$ such that

$$F(h_0; \theta_i) = \int_{-\infty}^{-(X\beta)_i} \varphi(t) \, dt, \tag{7}$$

where we use i to index the units.

We note that for a given h_0 this is a flexible model relating the Poisson parameter θ_i for an individual unit to the covariate values. Thus, given $(X\beta)_i$, θ_i is determined. It is convenient to choose $h_0 = 0$ so that

$$F(0; \theta_i) = \exp(-\theta_i) = \int_{-\infty}^{-(X\beta)_i} \varphi(t) \, dt.$$

We sample from the underlying standard normal distribution interval as follows for an observed count h^*, where the interval is indicated by brackets $\{\ \}$.

For an observed count in the first category we sample a value from the lower tail of the normal distribution

$$N\{-\infty, -(X\beta)_i\} \tag{8}$$

and for the remaining categories we sample from

$$N\{a, b\}, \qquad a = \Phi^{-1}\{F(h^*; \theta_i)\}, \quad b = \Phi^{-1}\{F(h^* - 1; \theta_i)\}. \tag{9}$$

The resulting sampled values, just as in the ordered probit model, are thus sampled from the specified standard normal distribution. Since this a novel characterization of the Poisson distribution there is no standard term for our procedure, and we shall refer to it simply as a Poisson latent normal transformation.

To impute any missing values on the original scale we use the inverse transformation (Goldstein *et al.*, 2009).

We note that our models for ordered categories consider the latent normal distribution to be defined by the observed categories, whereas the Poisson model is defined on $(0, \infty)$ and so assumes that the distribution exists beyond the categories observed. This suggests that we would expect to obtain different estimates. Since our data counts have a specified minimum count of 1 and a specified maximum possible count, the Poisson assumption may not be reasonable, and this is confirmed in our example. In the Poisson case, in some circumstances, we might consider fitting a truncated distribution that covers only the observed categories, or a particular subset of categories specified in advance. One example is the zero-truncated Poisson distribution where a zero category exists but cannot be observed, but this does not seem to be a reasonable assumption for our data.

4. A growth data example

Our application uses data that consist of 1000 subjects with serial weight measurements between birth and age 10 years together with adult body mass index BMI and plasma glucose level (millimoles per litre) measured on individuals at around the age of 30 years. During infancy (0–2 years) the maximum number of measurement occasions observed is 9 and during childhood (2–10 years) it is 14. There are 7459 repeated measurements over childhood and 1000 adult BMI and glucose measurements. The data set was made publicly available at the Third International Congress on Developmental Origins of Health and Disease (Gillman *et al.*, 2007).

As explained in Section 1, the joint modelling of the growth data and the adult measurements provides a flexible prediction system. This is the part of the model that is of principal interest. We also include the counts of the number of measurements as a further two, level 2, responses in infancy and in childhood. The principal purpose of this is to see whether the number of measurement occasions is related to characteristics of growth. Thus, we might suppose that children with atypical characteristics may also have atypical growth patterns and that this may be associated with the propensity to continue participation in the study. A strong relationship between the count of the number of measurements and such growth parameters may thus be indicative of non-random dropout. Strictly speaking, a smaller number of measurement occasions is not always indicative of dropout, although it is associated with it in the present data set. A more sensitive measure would be to count the number of measurements within more narrowly defined age groups, but this has not been done for the present data.

The responses in our model for prediction are as follows. At level 1 we have repeated measures of the log(weight) of children up to 10 years (y_{1ij}) with birth weight (x) in kilograms as a

covariate. Growth is modelled by using a basic cubic polynomial in age. At adulthood we have measures on the same individuals of log(glucose level) (y_{2j}) and log(BMI) (y_{3j}). We write the model as

$$
\left.
\begin{aligned}
y_{1ij} &= \beta_{0j} + \beta_{1j}\delta_{ij} + \beta_{2j}t_{ij}\delta_{ij} + \beta_3 t_{ij}^2 \delta_{ij} + \beta_4 t_{ij}^3 \delta_{ij} + \beta_{5j}t_{ij}(1-\delta_{ij}) + \beta_6 t_{ij}^2(1-\delta_{ij}) + \beta_8 x_j + e_{ij}, \\
\beta_{0j} &= \beta_0 + u_{0j}, \qquad \beta_{1j} = \beta_1 + u_{1j}, \qquad \beta_{2j} = \beta_2 + u_{2j}, \qquad \beta_{5j} = \beta_5 + u_{5j}, \\
&e_{ij} \sim N(0, \sigma_e^2), \\
&y_{2j} = \gamma_1 + u_{6j}, \\
&y_{3j} = \gamma_2 + u_{7j}, \\
&(u_{0j}, u_{1j}, u_{2j}, u_{5j}, u_{6j}, u_{7j})^{\mathrm{T}} \sim N(0, \Omega_u).
\end{aligned}
\right\}
\tag{10}
$$

Here, in terms of our general notation in equation (1), we have two level 2 response variables $y_j^{(2)} = (y_{2j}, y_{3j})$ and a single level 1 response variable $y_{1ij}^{(1)} = y_{1ij}$. The term δ_{ij}, in the child growth model, is an indicator variable taking the value 1 if the weight measurement is taken in infancy (0–2 years) and 0 if taken in childhood (2–10 years) so that the first line of model (10) fits two disjoint polynomial models: one for infancy and one for childhood. The two are linked through the level 2 random effects. This separation was specified by the original investigators who wished to consider the two age ranges separately. An alternative would have been to fit, for example, a regression spline to allow a smooth relationship across the age range, but we have not pursued this here. Age (t) in infancy is centred at 1.0 years and in childhood at 5.5 years. Note that Ω_u is of order 6 since the quadratic and cubic terms in model (10) are assumed not to have random coefficients.

The count data are introduced as follows. We have two further responses at level 2, namely a count of the number of measurement occasions in infancy (y_{4j}) and in childhood (y_{5j}). The random (intercept) effects for these (on the latent normal scale) at level 2 are respectively u_{8j} and u_{9j}. The only covariates are the intercepts for these responses.

For the Poisson model we have

$$
\mathrm{pr}(y_{4j} = h^*) = \int_a^b \phi(t)\, \mathrm{d}t
$$

where

$$
\mathrm{pr}(y_{4j} = 0) = \int_{-\infty}^{\gamma_4 + u_{8j}} \phi(t)\, \mathrm{d}t
$$

and

$$
\mathrm{pr}(y_{5j} = h^*) = \int_a^b \phi(t)\, \mathrm{d}t
$$

where

$$
\mathrm{pr}(y_{5j} = 0) = \int_{-\infty}^{\gamma_5 + u_{9j}} \phi(t)\, \mathrm{d}t
$$

where a and b are as defined in expressions (8) and (9).

For the ordered model with threshold parameters we have for infancy

$$
\mathrm{pr}(y_{4j} = h) = \int_{\alpha_{h-1} - (\gamma_4 + u_{8j})}^{\alpha_h - (\gamma_4 + u_{8j})} \phi(t)\, \mathrm{d}t, \qquad \alpha_0 = -\infty, \qquad \alpha_1 = 0, \qquad \alpha_{p_4} = \infty, \qquad h = 1, \ldots, p_4,
\tag{11}
$$

and for childhood

$$\mathrm{pr}(y_{5j}=h)=\int_{\alpha_{h-1}-(\gamma_5+u_{9j})}^{\alpha_h-(\gamma_5+u_{9j})}\phi(t)\,\mathrm{d}t, \qquad \alpha_0=-\infty, \qquad \alpha_1=0, \qquad \alpha_{p_5}=\infty, \qquad h=1,\ldots,p_5,$$

(12)

where p_4 and p_5 are the numbers of categories respectively for the numbers of occasions in infancy and childhood. For the smoothing model, equations (11) and (12) are modified accordingly. Thus the level 2 covariance matrix Ω_u has elements $\sigma_{u8}^2=\sigma_{u9}^2=1$ corresponding to the latent normal variables for the ordered responses. The Markov chain Monte Carlo step that samples the remaining elements of Ω_u first samples the variance terms followed by the correlations, which are then transformed to covariances.

5. Results

5.1. Latent normal models

Table 1 shows the estimates for the fixed parameters for three models. Model A treats the number of measurement occasions in infancy and childhood as ordered categories where each threshold parameter is estimated. Model B smooths the threshold parameters during infancy by using a third-order polynomial and the threshold parameters during childhood by using a fourth-order polynomial. Fitting a higher order polynomial in both cases does not change the other parameter estimates and the higher order polynomial coefficient estimates are smaller than their estimated standard errors. Model C fits a Poisson model. We comment further on these results below.

The chains for the threshold parameters were the least satisfactory in terms of how well they mixed. The burn-in period was chosen after observing the behaviour of these chains, especially for the extreme thresholds, using trial runs with different subsets of parameters. A value of 500 was found to provide a reasonable period for the chains to become stationary. Likewise, we found that 5000 iterations provided reasonable mixing and summary statistics in terms of means and standard deviations that were stable and did not change appreciably with further iterations, using similar trial runs. Further support for the values that were chosen is provided by the fact that models A and B, using different parameterizations, do provide similar estimates for the threshold parameters, albeit with different efficiencies (see below).

The level 2 covariance and correlation matrices are displayed in Table 2 where the variances are on the diagonal, and correlations below the diagonal. It will be seen that, whereas there are moderate correlations between the growth random effects, the correlations of growth random effects with adult measures are not large, so the usefulness of weight measures for prediction of the adult measures is limited. To illustrate the prediction procedure we consider predicting log(BMI) by using two childhood weight measurements, together with birth weight.

We write the 'raw' residuals for the two weight measurements at times t_{1j} and t_{2j} and the adult log(BMI) from model (10) as

$$\left.\begin{aligned}z_{1j}&=y_{11j}-(\beta_0+\beta_5t_{1j}+\beta_6t_{1j}^2+\beta_8x_j),\\z_{2j}&=y_{12j}-(\beta_0+\beta_5t_{2j}+\beta_6t_{2j}^2+\beta_8x_j),\\z_{3j}&=y_{2j}-\gamma_1.\end{aligned}\right\}$$

(13)

This defines a three-variate normal distribution that is given by

Table 1. Fixed effect and threshold estimates for the model defined by expressions (10)–(12)†

Coefficient	Model A	Model B	Model C
Intercept (β_0)	2.398 (0.013)	2.394 (0.013)	2.396 (0.013)
Infancy indicator (β_1)	−0.649 (0.003)	−0.649 (0.003)	−0.649 (0.003)
Age (infancy) (β_2)	0.027 (0.00051)	0.027 (0.000052)	0.027 (0.00052)
Age (infancy) squared (β_3)	−0.0027 (0.000034)	−0.0027 (0.000034)	−0.0027 (0.000034)
Age (infancy) cubed (β_4)	0.00017 (0.000055)	0.00017 (0.0000054)	0.00017 (0.0000056)
Age (childhood) squared (β_5)	0.010 (0.000060)	0.010 (0.0000062)	0.0100 (0.000060)
Age (childhood) cubed (β_6)	−0.000012 (0.0000012)	−0.000012 (0.0000012)	−0.000012 (0.0000012)
Birth weight (β_8)	0.172 (0.0033)	0.173 (0.0036)	0.172 (0.0035)
Level 2 intercept fixed effects			
log(glucose) (γ_1)	1.79 (0.0087)	1.79 (0.0087)	1.79 (0.0088)
log(BMI) (γ_2)	3.21 (0.0042)	3.21 (0.0042)	3.21 (0.0042)
Number of measures in infancy	1.09 (0.052)	1.08 (0.051)	1.62 (0.035)
Number of measures in childhood	2.89 (0.132)	2.76 (0.108)	2.95 (0.038)
Threshold parameters (infancy)			
α_2	0.93 (0.047)	0.92 (0.048)	
α_3	1.64 (0.058)	1.63 (0.058)	
α_4	2.22 (0.065)	2.23 (0.067)	
α_5	2.77 (0.079)	2.75 (0.076)	
α_6	3.23 (0.104)	3.20 (0.090)	
α_7	3.60 (0.143)	3.59 (0.128)	
α_8	4.11 (0.234)	3.94 (0.220)	
Threshold parameters (childhood)			
α_2	0.70 (0.141)	0.59 (0.123)	
α_3	1.47 (0.137)	1.28 (0.109)	
α_4	2.02 (0.135)	1.87 (0.109)	
α_5	2.56 (0.132)	2.38 (0.107)	
α_6	3.00 (0.130)	2.84 (0.106)	
α_7	3.44 (0.130)	3.25 (0.105)	
α_8	3.81 (0.134)	3.64 (0.105)	
α_9	4.19 (0.141)	4.01 (0.108)	
α_{10}	4.55 (0.145)	4.38 (0.115)	
α_{11}	4.90 (0.158)	4.73 (0.125)	
α_{12}	5.32 (0.176)	5.09 (0.140)	
α_{13}	5.66 (0.222)	5.43 (0.187)	
Polynomial smoother coefficients			
Infancy intercept		2.75 (0.076)	
Infancy linear		0.48 (0.032)	
Infancy quadratic		−0.035 (0.012)	
Infancy cubic		0.003 (0.003)	
Childhood intercept		3.45 (0.011)	
Childhood linear		0.39 (0.016)	
Childhood quadratic		−0.011 (0.006)	
Childhood cubic		0.002 (0.0008)	
Childhood quartic		−0.00013 (0.00022)	
Level 1 variance	0.0027 (0.000055)	0.0027 (0.000055)	0.0027 (0.000054)

†Burn-in, 500; sample size, 5000; number of level 1 units, 7459; number of level 2 units, 1000. Standard errors are given in parentheses. For the smoothed model the α-estimates are those derived from the smoothing polynomial at each iteration and the polynomial terms are centred on the mean category (numbered $1, \ldots, 8$ and $1, \ldots, 13$ for infancy and childhood respectively). Model A treats the number of measurement occasions in infancy and childhood as ordered categories where each threshold parameter is estimated. Model B smooths the threshold parameters by using a third-order polynomial for infancy and a fourth-order polynomial for childhood, and model C fits a Poisson model.

Table 2. Level 2 variances and correlations for model (10)–(12)†

Ordered categories

$$\Omega_u = \begin{pmatrix} 0.0053 \\ -0.74 & 0.0030 \\ 0.05 & -0.05 & 0.000010 \\ 0.51 & -0.60 & -0.43 & 0.0000028 \\ 0.10 & -0.15 & -0.08 & 0.21 & 0.0761 \\ 0.09 & -0.07 & -0.05 & 0.12 & 0.11 & 0.0172 \\ 0.008 & -0.010 & -0.014 & 0.03 & 0.05 & 0.04 & 1 \\ -0.003 & 0.004 & 0.004 & -0.008 & 0.02 & -0.02 & 0.014 & 1 \end{pmatrix}$$

Smoothed thresholds

$$\Omega_u = \begin{pmatrix} 0.0053 \\ -0.74 & 0.0031 \\ 0.08 & -0.08 & 0.000010 \\ 0.50 & -0.60 & -0.41 & 0.0000028 \\ 0.11 & -0.17 & -0.07 & 0.22 & 0.0761 \\ 0.09 & -0.07 & -0.04 & 0.11 & 0.11 & 0.0172 \\ 0.004 & -0.007 & -0.02 & 0.03 & 0.04 & 0.04 & 1 \\ -0.006 & 0.009 & 0.006 & -0.014 & 0.02 & -0.02 & 0.016 & 1 \end{pmatrix}$$

Poisson

$$\Omega_u = \begin{pmatrix} 0.0053 \\ -0.74 & 0.0031 \\ 0.10 & -0.10 & 0.000012 \\ 0.50 & -0.59 & -0.39 & 0.0000028 \\ 0.10 & -0.16 & -0.07 & 0.21 & 0.0762 \\ 0.08 & -0.06 & -0.04 & 0.11 & 0.10 & 0.0172 \\ 0.0001 & -0.003 & -0.02 & 0.03 & 0.04 & 0.04 & 1 \\ 0.002 & 0.0004 & 0.010 & -0.009 & 0.02 & -0.02 & 0.027 & 1 \end{pmatrix}$$

†The response variables are in the following order: overall intercept; infancy indicator; infancy slope; childhood slope; log(glucose); log(BMI); infancy count; childhood count.

$$\begin{pmatrix} z_{1j} \\ z_{2j} \\ z_{3j} \end{pmatrix} \sim N(0, \Omega_z),$$

$$\Omega_z = \begin{pmatrix} \sigma_{u0}^2 + 2t_{1j}\sigma_{u05} + t_{1j}^2\sigma_{u5}^2 + \sigma_e^2 \\ \sigma_{u0}^2 + (t_{1j}+t_{2j})\sigma_{u05} + t_{1j}t_{2j}\sigma_{u5}^2 & \sigma_{u0}^2 + 2t_{2j}\sigma_{u05} + t_{2j}^2\sigma_{u5}^2 + \sigma_e^2 \\ \sigma_{u06} + t_{1j}\sigma_{u56} & \sigma_{u06} + t_{2j}\sigma_{u56} & \sigma_{u6}^2 \end{pmatrix}.$$

From this normal distribution we can derive the expected value of any of the responses given the other two, which is just a linear (regression) predictor, which for z_{3j} has the form

$$\hat{z}_{3j} = \alpha_1 z_{1j} + \alpha_2 z_{2j}.$$

And the regression coefficients are derived from the above expression, namely

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \sigma_{u0}^2 + 2t_{1j}\sigma_{u05} + t_{1j}^2\sigma_{u5}^2 + \sigma_e^2 & \sigma_{u0}^2 + 2t_{2j}\sigma_{u05} + t_{2j}^2\sigma_{u5}^2 + \sigma_e^2 \\ \sigma_{u0}^2 + (t_{1j}+t_{2j})\sigma_{u05} + t_{1j}t_{2j}\sigma_{u5}^2 & \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{u06} + t_{1j}\sigma_{u56} \\ \sigma_{u06} + t_{2j}\sigma_{u56} \end{pmatrix}$$

where sample estimates, such as those given in Table 2, are substituted. The estimated covariance matrix of the coefficients is obtained from the inverse matrix given in the expression above and
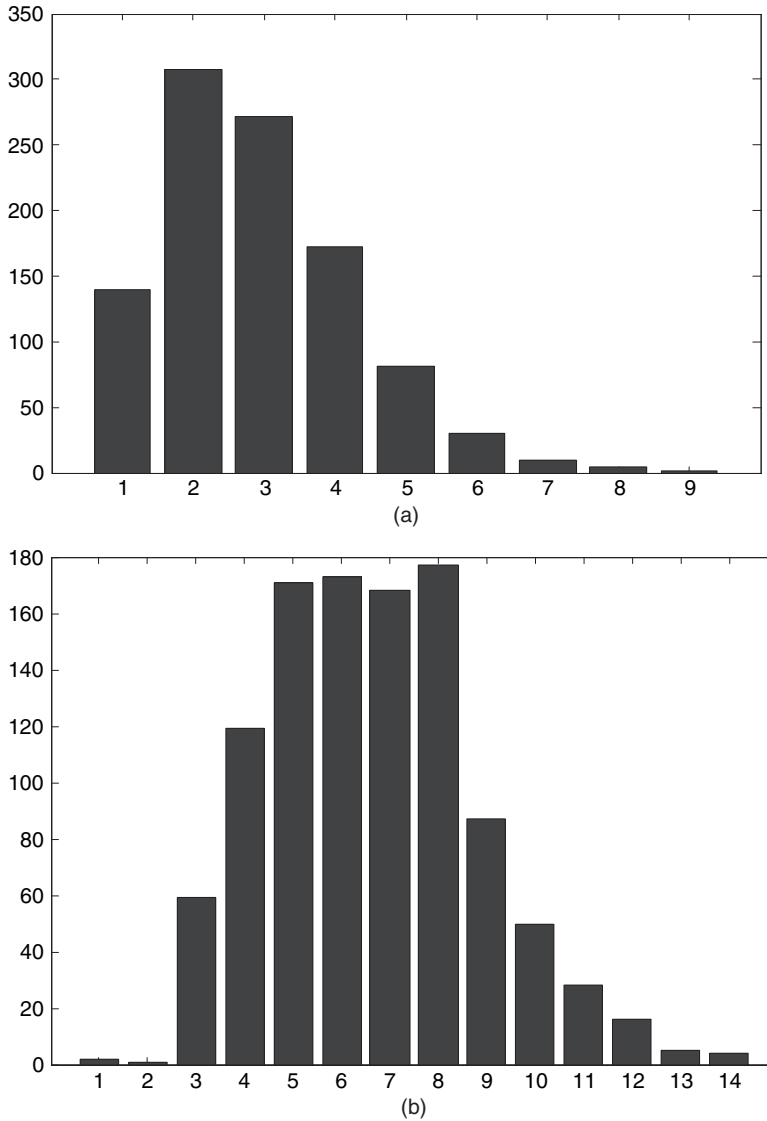
Fig. 1. Observed probability distributions for infancy and childhood number of measurement occasions (categories coded starting from 1; Poisson models were fitted): (a) infancy ($\chi^2 = 6.85$ (8 degrees of freedom), $P > 0.05$); (b) childhood ($\chi^2 = 39.1$ (13 degrees of freedom), $P < 0.001$)

this then enables us to place confidence intervals around the predicted values $\hat{z}_{3j}$. The predicted value of log(BMI) is given by combining $\hat{z}_{3j}$ with $\hat{\gamma}_1$ so that the predicted value of BMI becomes

$$\exp(\hat{z}_{3j} + \hat{\gamma}_1)$$

The count responses do not follow a Poisson distribution very closely, as is shown in Fig. 1; there is a non-significant χ^2 goodness-of-fit statistic for infancy with a significant value for childhood. This is also reflected in some of the estimates in Table 1 (see below) and the differences in the correlations in the last two rows of the correlation matrices in Table 2.

For all three models we obtain very similar estimates for the fixed effects, since the data are

Table 3. Fixed effect and threshold estimates for the model defined by expressions (10)–(13) with square-root transformations of the count responses†

Coefficient	Estimate
Intercept	2.396 (0.012)
Infancy indicator	−0.649 (0.003)
Age (infancy)	0.027 (0.00051)
Age (infancy) squared	−0.0027 (0.000032)
Age (infancy) cubed	0.00017 (0.0000054)
Age (childhood) squared	0.010 (0.000062)
Age (childhood) cubed	−0.000012 (0.0000012)
Birth weight	0.172 (0.0032)
Level 2 intercept fixed effects	
log(glucose)	1.79 (0.0087)
log(BMI)	3.21 (0.0041)
Square root of infancy count	1.24 (0.020)
Square root of childhood count	2.30 (0.016)
Level 1 variance	0.0027 (0.000056)

Level 2 covariance–correlation matrix

$$
\begin{pmatrix}
0.0053 & & & & & & & \\
-0.75 & 0.0030 & & & & & & \\
0.11 & -0.13 & 0.0000081 & & & & & \\
0.52 & -0.63 & -0.48 & 0.0000027 & & & & \\
0.11 & -0.17 & -0.09 & 0.22 & 0.075 & & & \\
0.09 & -0.07 & -0.06 & 0.12 & 0.11 & 0.017 & & \\
0.012 & -0.014 & -0.014 & 0.03 & 0.03 & 0.04 & 0.391 & \\
-0.006 & 0.009 & 0.005 & -0.013 & 0.019 & -0.024 & 0.005 & 0.247
\end{pmatrix}
$$

†Burn-in, 500; sample size, 5000; number of level 1 units, 7459; number of level 2 units, 1000. Standard errors are given in parentheses. The ordering of the random effects is as for Table 2.

complete with no missing values. An exception is for the Poisson model C for the level 2 infancy count and to a lesser extent for the childhood count. These estimates are larger since the Poisson distribution is defined for all non-negative values rather than being constrained by the range of observed values. For all the models we obtain reasonable Markov chain Monte Carlo chains for the parameters of interest. For the ordered unsmoothed model, however, some of the threshold parameter chains do not mix well. We also note, particularly, the smaller standard errors for the smoothing model compared with the unsmoothed ordered category model, especially for the childhood threshold parameters and the childhood count intercept at level 2. This illustrates the gain in efficiency from reducing the effective number of threshold parameters. Increasing the order of the polynomials leaves the estimates unchanged but tends to decrease precision so that care is needed to avoid oversmoothing since this will tend to decrease precision for other parameters that are associated with the variable being smoothed.

5.2. A square-root transformation model
We now compare these models with a model where we use a square-root transformation for the two count variables and then treat them as normal variables and fit a multivariate normal model. The results of this are displayed in Table 3.

We see that the fixed coefficient estimates and the level 2 intercept estimates for log(glucose) and log(BMI) are very close to those in Table 1 whereas the level 2 intercept estimates for the infancy and childhood count coefficients are somewhat different as a result of the square-root transformation.

6. Discussion

We have described the use of three different latent normal model formulations for count data. The simplest model, assuming a Poisson distribution, is straightforward to implement and can be combined with other normal or discrete responses in a multivariate model. The approach that we have adopted for a Poisson distribution is analogous to that used for a probit model for binary data and can be further extended to other discrete distributions. Where data do not follow, at least approximately, a Poisson distribution, we can treat the counts as an ordered classification, as we have done in the present application. In particular, we note that treating the counts as an ordered classification effectively conditions the latent normal distribution to be defined with respect to the actual range of observed counts. Where there are many categories we have shown how a smoothing formulation provides a good approximation with more efficient estimates.

The algorithm can be extended readily to further levels of nesting and cross-classifications. It is also relatively straightforward to incorporate linear constraints on sets of parameters (Goldstein *et al.*, 2009).

The data set that we have used has no missing values for the level 2 responses so the option that is chosen to model the count data will not affect the fixed parameter estimates for the remaining responses, as illustrated in Tables 1–3. Where data are missing and where, for example, the count responses are used in any kind of imputation procedure, particular assumptions, e.g. that the counts follow a Poisson distribution, will lead in general to somewhat different fixed and random-part estimates. The same will be true with the use of a square-root transformation where this does not lead to (approximate) normality.

The joint modelling of growth data and data on adults allows for the development of an efficient prediction system of adult values given a collection of growth measurements, and in principle this can be extended to several growth measures and further adult measures. We have also suggested that a model where the number of measurement occasions for different age ranges is jointly modelled with growth parameters may be useful for studying possible informative non-response patterns.

In the present example we only have birth weight as a covariate. In practice we may have other useful covariates such as social class and gender that can improve the prediction and can in principle be incorporated either as predictors or responses. In particular we may wish to use height as a 'covariate' since BMI is one of the adult measurements of interest, although it is not available in the present data set. In the case where height is not measured at the same times as weight, we may treat it as a further response that is modelled jointly with the other responses. When estimating the prediction we would then additionally condition on the observed height measurements.

The experimental software for fitting these models is written in MATLAB (Mathworks, 2007).

Acknowledgements

References

Browne, W. J. and Draper, D. (2006) A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayes. Anal.*, **1**, 473–514.

Gillman, M. W., Barker, D., Bier, D., Cagampang, F., Challis, J., Fall, C., Godfrey, K., Gluckman, P., Hanson, M., Kuh, D., Natanielsz, P., Nestel, P. and Thornburg, K. L. (2007) Meeting report on the 3rd International Congress on Developmental Origins of Health and Disease (DOHaD). *Ped. Res.*, **61**, 625–629.

Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edn. London: Arnold.

Goldstein, H. (2008) A general model for the analysis of discrete time multilevel survival data. To be published.

Goldstein, H., Carpenter, J., Kenward, M. and Levin, K. (2009) Multilevel Models with multivariate mixed response types. *Statist. Modllng*, to be published.

Goldstein, H., Steele, F., Rasbash, J. and Charlton, C. (2007) REALCOM: methodology for realistically complex multilevel modeling. Centre for Multilevel Modelling University of Bristol, Bristol. (Available from `http://www.cmm.bris.ac.uk/research/Realcom/index.shtml`.)

Hogan, J. W. and Laird N. M. (1997) Model based approaches to analysing incomplete longitudinal and failure time data. *Statist. Med.*, **16**, 259–272.

Mathworks (2007) *Matlab*. Natick: Mathworks. (Available from `http://www.mathworks.co.uk`.)

Tanner, J. M., Healy, M. J. R., Goldstein, H. and Cameron, N. (2000) *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*. London: Saunders.