Statistical Methods in Medical Research 2001; **10**: 409–427

Multilevel models for censored and latent responses

Sophia Rabe-Hesketh Department of Biostatistics and Computing, Institute of Psychiatry, King's College, London, UK, **Shuying Yang** MRC Biostatistics Unit, University of Cambridge, UK and **Andrew Pickles** School of Epidemiology and Health Science, The University of Manchester, UK

Multilevel models were originally developed to allow linear regression or ANOVA models to be applied to observations that are not mutually independent. This lack of independence commonly arises due to clustering of the units of observations into 'higher level units' such as patients in hospitals. In linear mixed models, the within-cluster correlations are modelled by including random effects in a linear model.

In this paper, we discuss generalizations of linear mixed models suitable for responses subject to systematic and random measurement error and interval censoring.

The first example uses data from two cross-sectional surveys of schoolchildren to investigate risk factors for early first experimentation with cigarettes. Here the recalled times of the children's first cigarette are likely to be subject to both systematic and random measurement errors as well as being interval censored. We describe multilevel models for interval censored survival times as special cases of generalized linear mixed models and discuss methods of estimating systematic recall bias.

The second example is a longitudinal study of mental health problems of patients nested in clinics. Here the outcome is measured by multiple questionnaires allowing the measurement errors to be modelled within a linear latent growth curve model. The resulting model is a multilevel structural equation model. We briefly discuss such models both as extensions of linear mixed models and as extensions of structural equation models. Several different model structures are examined.

An important goal of the paper is to place a number of methods that readers may have considered as being distinct within a single overall modelling framework.

1 Introduction

Multilevel models are models for grouped or hierarchical data, e.g. data on patients (level 1) who are nested in hospitals (level 2) which in turn may be nested in regions (level 3). The observations on different level 1 units belonging to the same level 2 (or higher level) unit cannot be assumed to be independent because they are likely to be influenced by the same (often unobserved) level 2 (or higher level) variables, e.g. hospital characteristics. Where it makes sense to consider these higher level units as drawn from some population then these influences can be modelled by random effects. If the response variable is continuous, the most commonly used multilevel model is a linear mixed model.

A three-level linear mixed model with levels $1, 2, 3$ indexed by k, j, i, can be written as

$$y_{ijk} = \boldsymbol{\beta}'\mathbf{x}_{ijk} + \mathbf{u}_{ij}^{(2)\prime}\mathbf{z}_{ijk}^{(2)} + \mathbf{u}_{i}^{(3)\prime}\mathbf{z}_{ijk}^{(3)} + \epsilon_{ijk} \quad (1)$$

Address for correspondence: S Rabe-Hesketh, Department of Biostatistics and Computing, Institute of Psychiatry, London SE5 8AF, UK. E-mail: spaksrh@iop.kcl.ac.uk

 0962-2802(01)SM267RA

where y_{ijk} is the response, $\mathbf{x}_{ijk}$ is a vector of q explanatory variables, $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients, $\epsilon_{ijk} \sim N(0, \sigma^2)$ is the residual (level 1) error term, $\mathbf{u}_{ij}^{(2)}$ and $\mathbf{u}_{i}^{(3)}$ are vectors of $m^{(2)}$ and $m^{(3)}$ random effects or coefficients that vary at levels 2 and 3, respectively, and $\mathbf{z}_{ijk}^{(2)}$ and $\mathbf{z}_{ijk}^{(3)}$ are the corresponding explanatory variables. The random coefficients are assumed to have multivariate-normal distributions with zero means, $\mathbf{u}_{ij}^{(2)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(2)})$ and $\mathbf{u}_{i}^{(3)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(3)})$. The explanatory variables $\mathbf{x}_{ijk}$, can vary at any level in the hierarchy, $\mathbf{z}_{ijk}^{(2)}$ must vary at level 1 and $\mathbf{z}_{ijk}^{(3)}$ must vary at level 2 or below. Each set of explanatory variables includes a constant and typically the $\mathbf{z}_{ijk}^{(3)}$ are subsets of the $\mathbf{z}_{ijk}^{(2)}$ which are subsets of the $\mathbf{x}_{ijk}$. The parameters of linear mixed models can be estimated by maximum likelihood (ML) or restricted ML in a large number of packages including MLwiN,[1] S-PLUS (lme), SAS (Proc Mixed) Stata (simple two-level models only), HLM[2] and VARCL.[3] Introductory books on the linear mixed model include Bryk and Raudenbush,[4] Longford,[5] Kreft and De Leeuw,[6] Snijders and Bosker[7] and Hox[8] (available free from the internet).

Two important extensions to linear mixed models are considered in this paper, generalized linear mixed models, and multilevel structural equation models. Section 2 introduces generalized linear mixed models, in particular models for discrete survival times, and illustrates their application to schoolchildren's retrospective reports of age of onset of smoking. Here multilevel models are required because children are nested in schools. Methods for handling bias in the recalled ages of onset are discussed. Section 3 introduces structural equation modelling of multilevel data, illustrated for data from a cluster-sampled two-phase longitudinal study of mental health problems. Here structural equation models allow a measurement model for the multiple fallible measures of mental health to be combined with a multilevel model for occasions nested in patients nested in clinics. ML estimation is used to handle data missing by design and due to attrition.

2 Generalized linear mixed models

Generalized linear mixed models can be used if the response variable is not conditionally normally distributed. A three-level generalized linear mixed model can be written as

$$g(E[Y_{ijk}|\mathbf{x}_{ijk}, \mathbf{z}_{ijk}^{(2)}, \mathbf{z}_{ijk}^{(3)}, \mathbf{u}_{ij}^{(2)}, \mathbf{u}_{i}^{(3)}]) = \boldsymbol{\beta}'\mathbf{x}_{ijk} + \mathbf{u}_{ij}^{(2)\prime}\mathbf{z}_{ijk}^{(2)} + \mathbf{u}_{i}^{(3)\prime}\mathbf{z}_{ijk}^{(3)} \quad (2)$$

where g is the link function and the conditional distribution of Y_{ijk} given the explanatory variables and random effects is from the exponential family. Familiar examples are logistic models for binomial responses and extensions to ordered and unordered multicategory responses, e.g. the proportional-odds and multinomial-logit models. The random effects are usually assumed to have multivariate-normal distributions but sometimes other distributions are assumed, e.g. a gamma distribution, because closed form solutions for the marginal or integrated likelihood are available. In general, the marginal likelihood cannot be evaluated exactly and parameters are estimated by approximate methods such as PQL/MQL[9] (used in MLwiN), quadrature[10–12] (used in Stata) or adaptive quadrature[13] (used in SAS), Markov Chain Monte Carlo[14] (used in BUGS and MLwiN). In economics, Geweke, Hajivassiliou and Keane[15] independently developed a method (known as GHK) of importance sampling for simulated moments that uses factorization to simulate conditional probabilities, making tractable problems

involving high dimensional integrals. Other simulation-based methods are reviewed in McCulloch.[16]

The fixed effects $\boldsymbol{\beta}$ are conditional effects, conditioning on the random effects. Whereas these effects are equal to the marginal effects with an identity link, this is not generally the case[17] (see also Diggle *et al.*[18]). An alternative approach to modelling clustered non-normal data is to use generalized estimating equations to model the marginal effects directly and assume a particular correlation structure for the responses although this does not correspond to any likelihood.

2.1 Multilevel discrete-time survival models

Survival times are frequently discrete or grouped. Reasons for this could be that the opportunities for the event occur at discrete times, as in the number of cycles to pregnancy, or that screening tests or surveys are carried out at discrete times, or simply that the times are reported in months or years, as is often the case with recalled onset data. If there are many ties among the survival times, methods for continuous survival time data such as Cox's regression are less satisfactory and discrete-time survival models should be used. Some of these models can be derived by assuming that the event actually occurred in continuous time but that we only know the interval in which the event occurred, i.e. treating the survival times as explicitly *interval censored*.

To introduce the ideas and notation, we will consider the example to be analysed later, where the outcomes are the ages when children reported having their first cigarette (ages of onset of experimenting with cigarettes). The survey responses provide six possible ages of onset (before age 11, age 11, 12, 13, 14, or age 15 and later) corresponding to the intervals $t_{k-1} \leq T < t_k$ where T is the unobserved continuous time of onset, $t_0 = 0$, $t_1 = 11$, $t_2 = 12$, $t_3 = 13$, $t_4 = 14$, $t_5 = 15$ and $t_6 = \infty$. As in the case of continuous survival times, times are right censored if the event has not occurred at the time of data collection. We consider two approaches to modelling discrete survival times, the proportional-odds model for ordinal data (with censoring) and models based on the discrete-time hazard.

2.1.1 Proportional-odds model for ordinal data

We can use the proportional-odds model for ordinal data[19] to model the discrete survival times. The only modification of the standard model is to allow for right censoring. The proportional-odds model assumes that the log of the odds that the event occurred before t_k is given by

$$\ln \frac{\Pr(T < t_k)}{1 - \Pr(T < t_k)} = \boldsymbol{\beta}'\mathbf{x} + \tau_k \tag{3}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients, $\mathbf{x}$ is a vector of explanatory variables and τ_k is a constant. (Note that the linear predictor of the proportional-odds model is normally defined as $-\boldsymbol{\beta}'\mathbf{x} + \tau_k$, but we reversed the sign of $\boldsymbol{\beta}$ so that, as with hazards-based models, positive coefficients imply earlier onset.) Therefore the probability that $T < t_k$ is

$$P_k = \Pr(T < t_k) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x} + \tau_k)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x} + \tau_k)} \tag{4}$$

and the probability that the survival time lies in the kth interval $t_{k-1} \leq T < t_k$ is

$$p_k = \Pr(t_{k-1} \leq T < t_k) = P_k - P_{k-1} \tag{5}$$

with $P_0 = 0$ and $P_6 = 1$. In the absence of censoring, this is the likelihood contribution of all observations whose survival times lie in the kth interval. For observations that are censored after the kth interval, the likelihood contribution is $1 - P_k$. The proportional-odds model has been used for continuous survival time data by Bennett,[20] and for discrete survival time data by Hedeker *et al*.[21]

The proportional-odds model can also be interpreted as a linear model for an underlying or latent continuous response y^*

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \epsilon \tag{6}$$

where ϵ has a logistic distribution. (If a standard-normal distribution is assumed for ϵ, the ordinal-probit model is obtained.) The event occurs in the kth interval if $\tau_{k-1} \leq y^* < \tau_k$, i.e.

$$\Pr(T < t_k) = \Pr(y^* < \tau_k) \tag{7}$$

The latent response y^* can therefore be thought of as a monotonic transformation of T so that $y^* = \tau_k$ corresponds to $T = t_k$. By constraining the threshold parameters τ_k to be equally spaced, the appropriateness of the linear regression model in (6) for the (untransformed) continuous time can be assessed.

2.1.2 Models based on the discrete-time hazard

The discrete-time hazard h_k for the kth interval is defined as the probability that the event occurs in the kth interval given that it has not already occurred

$$h_k = \frac{\Pr(t_{k-1} \leq T < t_k)}{\Pr(t_{k-1} \leq T)} \tag{8}$$

The likelihood contribution of someone whose survival time lies in the kth interval is

$$h_k \prod_{l=1}^{k-1}(1 - h_l) = \prod_{l=1}^{k} h_l^{y_l}(1 - h_l)^{(1-y_l)} \quad \text{with } y_k = 1 \tag{9}$$

where y_l is an indicator variable that is equal to 1 if the event occurred in the lth interval and equal to 0 otherwise, i.e. $y_l = 0$ when $l < k$ and $y_l = 1$ when $l = k$. The likelihood contribution of someone who was censored after the kth interval has the same form with

$$\prod_{l=1}^{k}(1 - h_l) = \prod_{l=1}^{k} h_l^{y_l}(1 - h_l)^{(1-y_l)} \quad \text{with } y_k = 0 \tag{10}$$

The likelihood contributions of both censored and non-censored observations are just the likelihood contributions of k-independent binary responses y_l, $l = 1, \ldots, k$ with Bernouilli probabilities h_l. Therefore, by expanding the data to k records per person and constructing the indicator variable y_l, discrete-time survival models can be fitted using standard software for binary responses. One possibility is to use logistic regression with a separate

constant for each interval

$$\log[h_l/(1-h_l)] = \boldsymbol{\beta}'\mathbf{x} + \tau_l \tag{11}$$

Note that this model is often referred to as a proportional-odds model. However, whereas proportionality here applies to the conditional odds of the event happening in an interval given that it has not already happened, proportionality in the previous section applied to the odds of the event happening in a given interval or earlier. Another term for this model is the continuation ratio logit model[22] or the logistic model for discrete-time survival data. To avoid confusion, we will refer to this model as the continuation ratio model. (See Jenkins[23] or Singer and Willett[24] for very good introductions to the model.)

If a Cox proportional-hazards model is assumed for the unobserved continuous survival times and the observed discrete survival times are treated as interval censored, it can be shown that the likelihood contributions are equal to those in (9) and (10) if a complementary log–log link is used for the discrete-time hazard,[25–27] i.e.

$$\log(-\log(1-h_l)) = \boldsymbol{\beta}'\mathbf{x} + \tau_l \tag{12}$$

Multilevel models for discrete-time survival data can be constructed by adding random effects to the linear predictors in (3), (11) and (12) as in (2). If a single log-gamma distributed random effect is used in the proportional-hazards model, closed form solutions exist for the log-likelihood[28] whereas generally, approximate methods must be used.[21] For the models based on the discrete-time hazard, standard mutlilevel software for binary responses can be used (e.g. Stata, MLwiN and SAS). Software for estimation of the proportional-odds model with random effects and censoring is less commonly available (MIXOR[11] can be used). Another variation of the models that we consider can be obtained by choosing a probit link function. Such models then potentially fall within the scope of structural equation models for categorical outcomes that are based on an underlying multivariate-normal latent covariance matrix (see Section 3.2). Estimation within such programs as Mplus[29] and Mx[30] then becomes possible but is probably only practical for small cluster sizes.

2.2 Analysis of age of onset of smoking data

We will now analyse data from two of a series of cross-sectional studies,[31,32] one from 1990 and the other from 1993, on the smoking behaviour of schoolchildren aged 11–15. Both studies followed similar two-stage sampling designs with schools as primary sampling units. The 1990 sample includes 3124 pupils from 125 schools and the 1993 sample includes 3140 different children in 110 different schools. The 5 years of classes sampled within each survey and the 3-year interval between surveys resulted in some age cohorts being sampled twice (e.g. the 14-year-olds in 1993 are in the same cohort as the 11-year-olds in 1990). The sampling fraction for schools has been assumed to be sufficiently low that the possibility of schools appearing in both the 1990 and 1993 samples could be ignored. The children were asked whether they had ever smoked a cigarette, and if so, how old they were the first time they smoked. In 2% of observations the child did not remember the age of onset (left censoring) and these observations were dropped. When the age of onset was equal to the current age, the observation was treated as right censored to avoid any potential biases due to children being surveyed after different lengths of time since their last birthdays.

The available explanatory variables that we consider as possible influences on age of first experimentation with smoking are sex, cohort (modelled as a linear effect), a sex by cohort interaction and a binary variable for the presence of a smoking parent at home. A random intercept for school is included to account for the cluster sampling design. The linear predictor for the jth pupil in the ith school is therefore

$$\boldsymbol{\beta}'\mathbf{x}_{ij} + u_{0i}^{(2)} + \tau_k \tag{13}$$

All models were fitted using the Stata program gllamm (formerly known as gllamm6[12,33]) which maximizes the marginal log-likelihood approximated by Gaussian quadrature (20 quadrature points were used). The continuation ratio and proportional-hazards models were fitted by expanding the data as outlined in Section 2.1.2 and fitting mixed logistic regression and mixed complementary log–log models, respectively. The mixed proportional-odds model was fitted by treating the responses of children of different current ages as distinct ordinal responses with different numbers of categories (see Table 1). For example, those who were aged 11 when surveyed have an ordinal response with two possible categories and those who were aged 12 have a response with three categories, etc. The thresholds τ_k were constrained equal across responses. The parameter estimates are given in the first three columns of Table 2. All three models lead to essentially the same conclusions. Females are less at risk than males if no parent is smoking at home. A parent smoking at home increases the risk of smoking for boys and this effect is even greater for girls. There is a significant linear effect of cohort for girls with the risk of having a first cigarette earlier increasing over time. There is no significant cohort effect for boys. There is significant heterogeneity between schools in the ages of onset (likelihood ratio test, $p < 0.001$ in all models.

One problem with these three models is that they assume that the recalled ages of onset are reliable. Accounting for measurement error in age-of-onset data has received rather

Table 1 Ages, possible times of onset and associated probabilities

Age	11	12	13	14	15
Proportional-odds model	$T<11$ (P_1)	$T<11$ (P_1)	$T<11$ (P_1)	$T<11$ (P_1)	$T<11$ (P_1)
	$11\leq T$ $(1-P_1)$	$11\leq T<12$ (P_2-P_1)	$11\leq T<12$ (P_2-P_1)	$11\leq T<12$ (P_2-P_1)	$11\leq T<12$ (P_2-P_1)
	– –	$12\leq T$ $(1-P_2)$	$12\leq T<13$ (P_3-P_2)	$12\leq T<13$ (P_3-P_2)	$12\leq T<13$ (P_3-P_2)
	– – – –	– – – –	$13\leq T$ $(1-P_3)$ – –	$13\leq T<14$ (P_4-P_3) $14\leq T$ $(1-P_4)$	$13\leq T<14$ (P_4-P_3) $14\leq T<15$ (P_5-P_4)
	– –	– –	– –	– –	$15\leq T$ $(1-P_5)$
Current-status model	$T<11$ (P_1)	$T<12$ (P_2)	$T<13$ (P_3)	$T<14$ (P_4)	$T<15$ (P_5)
	$11\leq T$ $(1-P_1)$	$12\leq T$ $(1-P_2)$	$13\leq T$ $(1-P_3)$	$14\leq T$ $(1-P_4)$	$15\leq T$ $(1-P_5)$

Table 2 Parameter estimates for the multilevel proportional-hazards, continuation-ratio and proportional-odds models estimated in gllamm using 20 quadrature points

Parameters	Continuation ratio		Proportional hazards		Proportional odds		Current status		Telescoping	
	Estimate	Parameters	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Female pupil	−0.336	0.118	−0.304	0.108	−0.451	0.140	−0.322	0.083	−0.305	0.151
Parent smokes	0.309	0.066	0.281	0.060	0.374	0.077	0.352	0.083	0.377	0.079
Female pupil × parent smokes interaction	0.288	0.094	0.264	0.086	0.327	0.110	0.384	0.118	0.358	0.113
Cohort effect										
Boys	0.020	0.019	0.018	0.017	0.016	0.021	0.037	0.028	0.034	0.027
Girls	0.076	0.019	0.070	0.017	0.082	0.021	0.086	0.028	0.082	0.027
Telescoping										
Boys									−0.098	0.036
Girls									0.037	0.037
Age thresholds										
11	−2.172	0.100	−2.223	0.093	−2.178	0.114	−2.428	0.199	−2.373	0.184
12	−2.414	0.102	−2.449	0.094	−1.524	0.110	−1.720	0.174	−1.678	0.159
13	−1.907	0.098	−1.980	0.090	−0.887	0.107	−0.910	0.152	−1.012	0.139
14	−1.380	0.098	−1.508	0.088	−0.221	0.104	−0.299	0.134	−0.312	0.122
15	−1.227	0.116	−1.380	0.103	0.323	0.106	0.189	0.115	0.280	0.115
School variance	0.067	0.019	0.057	0.016	0.091	0.025	0.095	0.028	0.087	0.026
Log-likelihood	−6225.5		−6225.8		−6223.4		−3487.7		−5948.8	

little attention in the literature. We consider two alternative approaches. The first approach is to discard the timing element of the children's responses and simply model their current status (ever experimented) as a function of their current age, using a simple logistic regression model with the current smoking status indicator as the response variable, as indicated in Table 1. This gives the results in column 4 of Table 2 which are not very different from those for the proportional-odds model. Another approach is to model recall bias directly. It has been suggested that recall errors are characterized by an apparent shifting of events from the more distant past towards the time at which data collection is made.[34,35] This 'telescoping' could arise from an internal compression of the time scale so that an event that occurred a time t ago is reported as occurring a time γt ago with $0 < \gamma < 1$. Telescoping could also result from heteroscedastic measurement error in which the error variance increases with the lag between the event and the time of recollection even when the errors are symmetrically distributed. This is because more events from the distant past, that are typically recalled with larger errors, are shifted into the recent past than events in the recent past, that are typically recalled with smaller error, are shifted back into the distant past.[36] While Pickles *et al.*[36] develop models to distinguish between these processes, here we will only consider systematic telescoping.

In the proportional-odds model, we assume that the log odds that the recalled age of onset is before a given age t_k decreases linearly with the time that has passed since that age, $a_{ij} - t_k$, where a_{ij} is the child's current age

$$\ln \frac{\Pr(T_{ij} < t_k)}{1 - \Pr(T_{ij} < t_k)} = \boldsymbol{\beta}'\mathbf{x}_{ij} + u_{0i}^{(2)} - \boldsymbol{\alpha}'\mathbf{w}_{ij}(a_{ij} - t_k) + \tau_k \tag{14}$$

$\boldsymbol{\alpha}$ is a vector of coefficients and $\mathbf{w}_{ij}$ is a vector of explanatory variables that may predict the degree of telescoping (positive coefficients imply a compression of the time scale). If the proportional-odds model is interpreted as a latent response model, telescoping corresponds to allowing the thresholds to depend on the time-lag, i.e. the thresholds are

$$\tau_k - \boldsymbol{\alpha}'\mathbf{w}(a_{ij} - t_k) \tag{15}$$

Here we assume that the degree of telescoping depends on sex only, giving the parameter estimates in the last column of Table 2. While there is no significant telescoping for girls, the boys tend to stretch the time scale (rather than compress it), perhaps 'showing off' with having experimented earlier than they actually did. Note that separate identification of telescoping and cohort effects is possible here because some of the cohorts of children are represented in both surveys at different 'current' ages and therefore with different time lags.

3 Multilevel structural equation models

Multilevel structural equation models for continuous data can be approached in two ways. The first, probably more accessible to medical statisticians, is to extend a familiar multilevel model to tackle a latent variable problem. The second is to extend the simple latent variable model to a multilevel context. We shall describe both for the case of a multilevel latent growth curve model.

3.1 Generalizing growth curve models to multilevels and multivariate measurements

We begin by considering extensions to the multilevel linear growth model that can be used when the data structure is hierarchical and any of the observed (response or explanatory) variables are subject to measurement error. In so doing we generalize the concept of a random effect to that of a latent variable. In the optimistic terminology of psychological measurement theory, the latent variables represent the 'true' values of underlying constructs that are measured by error-prone observed variables.

If there are several fallible measurements y_{ijs} of the 'true' outcome T_{ij} for subject i at occasion j, e.g. arising from different 'instruments' s, a possible measurement model for y_{ijs} is

$$y_{ijs} = \alpha_s + \lambda_s T_{ij} + \epsilon_{ijs} \tag{16}$$

which has a factor analysis form where α_s is the instrument-specific bias, λ_s is an instrument-specific scaling factor or 'factor loading' and ϵ_{ijs} is measurement error. For identification, $\alpha_1 = 0$ and $\lambda_1 = 1$ are imposed, fixing the location and scale of measurement of T_{ij} to be that of instrument 1. The variance of the measurement errors ϵ_{ijs} are assumed to differ between the instruments.

We can then assume a growth curve model for the 'true' outcome

$$T_{ij} = \beta_0 + \beta_1 t_{ij} + u_{0i}^{(3)} + u_{1i}^{(3)} t_{ij} + u_{0ij}^{(2)} \tag{17}$$

where t_{ij} is time, β_0 is the mean intercept, $u_{0i}^{(3)}$ is subject i's random deviation from the mean intercept, β_1 is the mean slope, $u_{1i}^{(3)}$ is the deviation of subject i's slope from the mean slope and $u_{0ij}^{(2)}$ is a residual error term. Note that the level 1 units in this model are the measurements of an individual at a given occasion using multiple scales, so that occasions are at level 2 and subjects at level 3. Substituting (17) into (16) gives

$$y_{ijs} = \alpha_s + \lambda_s\beta_0 + \lambda_s\beta_1 t_{ij} + \lambda_s u_{0i}^{(3)} + \lambda_s t_{ij} u_{1i}^{(3)} + \lambda_s u_{0ij}^{(2)} + \epsilon_{ijs} \tag{18}$$

If the longitudinal data are balanced, with $t_{ij} = t_j$, this 'latent growth curve model' can also be viewed as an ordinary structural equation model (see next section). The model becomes a multilevel structural equation model only if higher level random effects are introduced. Here we outline our own approach[37,38] (Yang and Pickles, submitted) of estimating such multilevel structural equation models by reparameterizing equations of the form of (18) as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{l=2}^{L} \mathbf{H}_l \mathbf{u}_i^{(l)} + \epsilon_i \tag{19}$$

where $\mathbf{y}_i$ is the vector of responses y_{ijs} for the ith cluster and $\mathbf{X}_i$ represents the design matrix for all the fixed effect explanatory variables. The $\mathbf{u}_i^{(l)}$ are vectors of random effects or latent variables that vary at level l ($l = 2, \ldots, L$), the random measurement-specific error ϵ_i occupying level 1. The $\mathbf{u}_i^{(l)}$ are assumed to be normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}^{(l)}$. The residuals ϵ_i are normally distributed with mean $\mathbf{0}$ and diagonal covariance matrix $\boldsymbol{\Sigma}$.

The fixed effects parameters $\boldsymbol{\beta}$ in (19) are simple functions of the intercepts and regression coefficients of the growth curve and the intercepts and factor loadings of the measurement model. The matrix $\mathbf{H}_l$ multiplying the latent variables in (19) has elements that are products of regression coefficients or factor loadings and a design matrix. Typically some standard restrictions to the full set of parameters are required to achieve identifiability. Exactly what parameters and functions of parameters are estimated depends upon the choice of restrictions and the corresponding design matrix. Computation proceeds treating each cluster i as a single block, unbalanced data being easily accommodated by allowing the block dimensions, and hence the number of rows of $\mathbf{X}_i$ and $\mathbf{H}_l$ to vary with i.

If the mth element of $\mathbf{u}_i^{(l)}$, $u_{mi}^{(l)}$, represents the 'true' value of some imperfectly measured response, then the corresponding (mth) column of $\mathbf{H}_l$ contains the factor loadings. If the vector $\mathbf{u}_i^{(l)}$ is a set of random coefficients (e.g. intercept and slope), then the rows of $\mathbf{H}_l$ are the vectors of explanatory variables $\mathbf{z}_{ijs}^{(l)\prime}$ (typically $\mathbf{H}_l$ contains some of the columns of $\mathbf{X}$).

Although maximization of the full log-likelihood is possible, an iterative three-step ML procedure has proved to be much faster (for ordinary multilevel models a two-step iterative least squares estimation method has been used by Goldstein[39,40] where fixed effects are estimated in the same way). In the first step, we estimate the regression coefficients $\boldsymbol{\beta}$ using least squares, assuming that the variance covariance matrix of $\mathbf{y}_i$ is known. The second step estimates any unknown parameters in $\mathbf{H}_l$, based on the estimated $\boldsymbol{\beta}$, and the initial $\boldsymbol{\Sigma}^{(l)}$ and $\boldsymbol{\Sigma}$. In the third step, the variance covariance matrices $\boldsymbol{\Sigma}^{(l)}$ and $\boldsymbol{\Sigma}$ are estimated based on the parameter estimates from the previous steps. The second and third steps are achieved by directly maximizing the multivariate-normal log-likelihood. These three steps are iterated until convergence to the ML estimates. Finally, standard errors are obtained by inverting a numerically derived Hessian for the full log-likelihood estimated at the ML parameter estimates. This separation of point and precision estimation can be used to advantage, by exploiting the opportunity to make a reparameterization between them. This is helpful where the estimated parameters of $\boldsymbol{\beta}$ and $\mathbf{H}_l$ involve functions of the fundamental parameters rather than being the fundamental parameters themselves, or in other cases where the theoretically interesting parameterization for which we want estimates and standard errors has inferior optimization properties compared to a parameterization for which estimation is more tractable.

This algorithm has been programmed in S-PLUS for Windows, using the function nlmin() (StatSci 1993) for the log-likelihood maximization of steps 2 and 3 above. For increased speed, the likelihood is evaluated by a linked C-program. The program is available from the authors. Note that the model can also be estimated in gllamm (see e.g. Rabe-Hesketh *et al.*[41]) but the current quadrature approximation is unlikely to work as well for normally distributed responses as the iterative method outlined above.

3.2 Generalizing structural equation models

Here we derive our multilevel growth curve model starting from a standard structural equation model. Books on structural equation models include Dunn *et al.*[42] and Bollen[43] and software to fit these models include EQS,[44] Mplus,[29] Amos,[45] Mx[30] and LISREL.[46] Commonly such models consist of two parts, a measurement model and a structural model. In the measurement part, the observed variables are regressed on the latent

variables, though if required exogeneous explanatory variables measured without error can also be included among the predictors. In the structural model latent variables are regressed on other latent and explanatory variables.

Using the notation of Muthén,[47] the measurement part of the model can be written as

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{x}_i + \boldsymbol{\epsilon}_i \tag{20}$$

where $\mathbf{y}_i$ is a p-dimensional response vector, $\boldsymbol{\nu}$ is a p-dimensional vector of constants or intercepts, $\boldsymbol{\eta}_i$ is an m-dimensional vector of latent variables, $\boldsymbol{\Lambda}$ is an $p \times m$ matrix of regression coefficients (factor loadings), $\mathbf{x}_i$ is a vector of $q-1$ explanatory variables, $\mathbf{K}$ is an $m \times (q-1)$ matrix of regression coefficients and $\boldsymbol{\epsilon}_i$ is a p-dimensional vector of residuals.

This measurement part can be used to define linear growth curve models or other two-level models for balanced data (see e.g. Dunn *et al.*;[42] Little *et al.*[48]).

The structural part of the model can be written as

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{w}_i + \boldsymbol{\zeta}_i \tag{21}$$

where $\boldsymbol{\alpha}$ is an m-dimensional parameter vector, $\mathbf{B}$ is an $m \times m$ matrix of regression coefficients for regressions of latent variables on other latent variables (the diagonals are zero), $\boldsymbol{\Gamma}$ is an $m \times (r-1)$ matrix of regression coefficients for regressions on the observed variables $\mathbf{w}_i$ and $\boldsymbol{\zeta}_i$ is an m-dimensional vector of residuals.

Typically, $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ are not of interest in structural equation modelling and estimation is based only on the sample covariance matrix. However, for growth curve modelling, 'mean structure analysis' is used where $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ are estimated together with the other parameters using both the sample covariance matrix and the sample means. If there are missing data, the problem can be set up as a multiple group analysis where the data are split into groups with different patterns of missing data and all parameters are constrained equal between groups.[49] This method is not practical when there are many different patterns of missing data or when the number of observations for some patterns are small. Several packages (e.g. Mplus, Amos and Mx) employ full information ML estimation for data with missing values. (Here estimation is based on the whole data matrix.)

Substituting the structural equation into the measurement equation gives

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\alpha} + \mathbf{K}\mathbf{x}_i + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{w}_i + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_i \tag{22}$$

The correspondence between this general structural equation model and the parameterization in equation (19) can be seen by setting $L = 2$, $\mathbf{u}_i^{(2)} = \boldsymbol{\zeta}_i$ and $\mathbf{H}_2 = \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}$. The first four terms correspond to the fixed part of (19). Whereas the framework in (19) implies that the values of the explanatory variables will differ between the observations represented in $\mathbf{y}_i$ and the coefficients will not, the structural equation approach assumes constant $\mathbf{x}_i$ and $\mathbf{w}_i$ with observation (or response)-specific coefficients. Use of dummy variables in (19) and parameter constraints in (22) allows both frameworks to accommodate both scenarios.

Multilevel structural equation models can be defined by allowing the observed variables $\mathbf{x}_i$ and $\mathbf{w}_i$ and the latent variables $\boldsymbol{\eta}_i$ to vary at different levels. Consistent with the previous section, we will refer to the variables of a multivariate response as level 1 units so that $\boldsymbol{\zeta}_i$ is a level 2 latent variable. Most attempts to define and estimate multilevel structural equation models have been restricted to three-level models and consist of

specifying separate structural equation models for the between- and the within-level 3 covariance matrix,[50–53] although McDonald and Goldstein define more general multilevel structural equation models.[54,55] One approach, based on splitting the sample covariance into a between and within matrix[53,56] yields ML parameter estimates only in the balanced case.[57] Muthén's Mplus program extends this method to the unbalanced case using the MUML estimator (Muthén's ML-based estimator)[57,58] which is not an ML estimator. When data are missing, the program analyses complete cases only, but a multiple group analysis can be used to accommodate different missing data patterns.

In practice direct ML estimation is now possible in both the balanced and unbalanced case,[50] e.g. using an EM algorithm.[59,60] Alternatively, ML estimation in the general unbalanced and missing at random case is possible using our approach described in the previous section.

3.3 Latent growth model for clustered two-phase longitudinal survey data

The Psychological Problems in General Health Care Study[61,62] involved a multicentre two-phase design. All subjects were screened using a 12 item 'General Health Questionnaire' (GHQ) at the first phase (time 1). Patients were selected to enter into a second phase (interview) based on their GHQ-12 screen scores. The individuals in the second phase were followed up by additional surveys using a novel 34-item version soon after the screening test (time 2), and the 28-item version (GHQ-28) at 3 months (time 3) and then 1 year later (time 4). The interest is to study the changes of GHQ over time in the population of people in primary care.

The 34-item version includes all the items of the GHQ-12 and GHQ-28, but the GHQ-28 includes only six of the GHQ-12 items. All items were scored on four-point scales (the greater the score, the more distressed the patient). Normally, the total of all the items of the questionnaire would be used as a measure of distress. However, such totals would not enable analysis of changes over time since different versions of the GHQ were used at different occasions. We therefore grouped the items into three subscales so that each subscale is available on multiple occasions. The subscales were the six items measured at all four occasions (subscale 1), the remaining six items of the 12-item GHQ (subscale 2) and the 22 additional items in GHQ-28 (subscale 3). Thus in this study, all subjects had subscales 1 and 2 measured at time 1, but only those who entered into the second phase had (apart from missing data) subscales 1, 2 and 3 at time 2 and, subscales 1 and 3 at times 3 and 4.

The data we consider in this paper are from a total of 2094 patients from the 42 participating Paris clinics. The size of the clinics varied from 2 to 232. Of these, 382 patients met the selection criteria and were involved in the second phase of the study. Only 187 patients had all measurements at all occasions. A natural logarithm transformation was applied to the subscale totals to adjust for skewness.

We will use indices i for clinics, j for subjects, $r = 1, 2, 3, 4$, for occasions and $s = 1, 2, 3$ for subscales. The different subscale scores at a particular time r can be assumed to measure the same latent trait, 'psychological distress' at time r but on different scales and with differing precision. An appropriate measurement model may be

$$\begin{aligned} y_{ijrs} &= \alpha_{rs} + \lambda_s(\nu_r + \eta_{ijr}) + \epsilon_{ijrs} \\ &= \nu_{rs} + \lambda_s \eta_{ijr} + \epsilon_{ijrs} \end{aligned} \tag{23}$$

where α_{rs} is an occasion and subscale-specific bias, ν_r is the mean true psychological distress at time r, $\nu_{rs} = \alpha_{rs} + \lambda_s \nu_r$, η_{ijr} is the deviation of subject ijth's true distress from the mean at time r, λ_s is a factor loading and $\mathrm{var}(\epsilon_{ijrs}) = \sigma_s^2$ is the measurement error variance. The factor loadings and measurement error variances reflect both the scale and reliability of the instrument and are assumed to be constant over time. In order to identify the scale of the latent variable, the factor loading for the first subscale is fixed at 1. We assume that the deviation of subject ij's true psychological distress from the population mean can be modelled by a growth curve

$$\eta_{ijr} = u_{0ij}^{(3)} + u_{1ij}^{(3)} t_r + u_{0i}^{(4)} + u_{1i}^{(4)} t_r + u_{0ijr}^{(2)} \tag{24}$$

where the level 3, subject-specific random effects, $u_{0ij}^{(3)}, u_{1ij}^{(3)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(3)})$ are independent of the level 4, clinic-specific random effects, $u_{0i}^{(4)}, u_{1i}^{(4)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(4)})$, which are independent of the level 2, occasion-specific random effects $u_{0ijr}^{(2)} \sim N(0, \psi)$ and t_r is equal to 0 at occasions 1 and 2, 1 at occasion 3 and 4 at occasion 3.

The full model is

$$y_{ijrs} = \nu_{rs} + \lambda_s \left(u_{0ij}^{(3)} + u_{1ij}^{(3)} t_r + u_{0i}^{(4)} + u_{1i}^{(4)} t_r + u_{0ijr}^{(2)} \right) + \epsilon_{ijrs} \tag{25}$$

By treating the realizations of $u_{0ijr}^{(2)}$, $r = 1, \ldots, 4$ as four separate (uncorrelated) latent variables at the subject level, as shown in path diagram form in Figure 1, the model without the level 4, clinic-specific random effects can be written as a standard structural equation model of the form of (20) and (21) and can be fitted using any structural equation packages that allows estimation with missing values. In the path diagram, boxes represent observed variables and circles represent latent variables. Arrows represent regressions and their labels represent the corresponding regression coefficients. Error terms of regressions are indicated by arrows pointing to the dependent variable. Double-headed arrows represent correlations. We have put a frame around the observed and latent variables that are clinic specific (subscript i) and another frame around all observed and latent variables that additionally vary between patients within clinics (subscript ij). The same model is written in the notation of equation (19) in the Appendix for a subject who has complete data.

The model parameters can be estimated by standard ML since the structured missing data arising from the two-phase design are missing at random.[63] Mplus was used to fit three different models without the level 4 random effects; the full model (Model 3), a model with no random slope (Model 1) and a model with no occasion-specific errors (Model 2). The results are given in Table 3. Using likelihood ratio tests, the occasion-specific error term is highly significant ($p < 0.001$) and the p-value for the random slope is 0.025.

The set-up to apply the direct ML approach described at the end of Section 3.1 is given in an Appendix. It gave essentially identical results as those from Mplus for the models excluding the clinic-level random effects. However, unlike Mplus, this approach allows the full model (Model 4) with clinic-specific random effects at level 4 to be fitted by ML. (Mplus cannot be used to estimate the parameters of the four-level model by ML because the number of patients per clinic varies between clinics and because of missing data, although it may be possible to handle the eight patterns of missing data using multiple group analysis.)

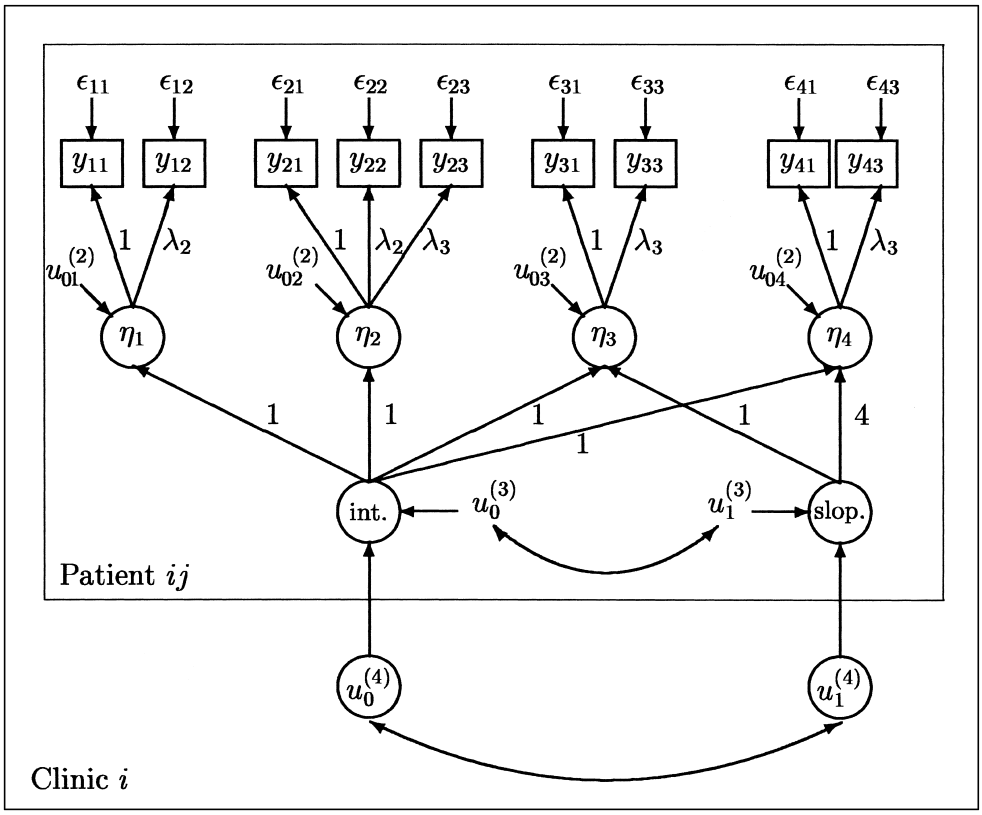

Figure 1 Path diagram of multilevel structural equation model for GHQ data

The results for this model are shown in the last two columns of Table 3. These results indicate a significant intercept variation among clinics. A further model, that is not shown, allowed for an additional clinic-level (level 4) random slope, but this proved to be non-significant.

4 Discussion

Approaching interval censored survival as a discretely observed latent continuous survival time provided scope for interesting model developments in two main ways. Firstly, the direct modelling of the survival time itself in the proportional-odds model readily accomodates a number of processes of scientific interest that are more difficult to represent in hazard model specifications. Our example introduced one such process, namely measurement error in time. Of course, standard hazards model specifications have the great advantage that they can be linked to parametric stochastic model theory, but the majority of medical and social applications do not exploit this link, preferring to fit non-parametric baseline hazards. It is arguable that comparable flexibility can be achieved

Table 3 Parameter estimates for multilevel structural equation models using Mplus and direct ML estimate method

Parameters	Model 1		Model 2		Model 3		Model 4	
	Estimate	Parameters	Estimate	SE	Estimate	SE	Estimate	SE
Subscale means (intercepts)								
ν_{11}	1.967	0.009	1.967	0.009	1.967	0.009	1.950	0.016
ν_{12}	1.909	0.010	1.909	0.010	1.909	0.010	1.888	0.020
ν_{21}	1.945	0.021	1.855	0.018	1.935	0.021	1.924	0.025
ν_{22}	1.910	0.024	1.794	0.018	1.897	0.024	1.883	0.029
ν_{23}	2.839	0.029	2.706	0.026	2.823	0.029	2.806	0.036
ν_{31}	1.751	0.021	1.684	0.017	1.746	0.021	1.735	0.024
ν_{33}	2.535	0.029	2.438	0.026	2.528	0.028	2.510	0.035
ν_{41}	1.693	0.024	1.692	0.026	1.701	0.025	1.692	0.028
ν_{43}	2.459	0.033	2.462	0.040	2.471	0.035	2.456	0.041
Factor loadings								
λ_2	1.245	0.031	1.260	0.034	1.248	0.031	1.250	0.031
λ_3	1.580	0.044	1.527	0.055	1.577	0.044	1.583	0.040
Clinic-level variance								
$\mathbf{\Sigma}^{(4)}$ – intercept variance							0.006	0.002
Subject level (co)variances								
$\mathbf{\Sigma}_{11}^{(3)}$ – intercept variance	0.059	0.005	0.100	0.005	0.064	0.005	0.058	0.005
$\mathbf{\Sigma}_{21}^{(3)}$ – intercept slope covariance			−0.010	0.002	−0.002	0.002	−0.003	0.001
$\mathbf{\Sigma}_{22}^{(3)}$ – slope variance			0.007	0.0001	0.002	0.001	0.002	0.001
Occasion-level variance								
ψ – intercept variance	0.054	0.003			0.052	0.004	0.048	0.003
Subscale error variances								
σ_1^2	0.051	0.003	0.067	0.003	0.051	0.003	0.052	0.002
σ_2^2	0.044	0.004	0.058	0.003	0.058	0.003	0.044	0.004
σ_3^2	0.051	0.006	0.144	0.009	0.051	0.006	0.050	0.006
Log-likelihood								
	−2340.465		−2643.433		−2336.752		−2311.20	

when modelling survival time itself and that for analysing commonly experienced outcomes, such as developmental milestones, that these are at least as natural a parameterization as hazards models. Secondly, when time is considered as a continous latent response it is straightforward to construct and estimate models for multivariate responses based on standard multivariate methods such as structural equation modelling, whether these be multivariate survival times, such as in twins studies[64] or a combination of survival times and other outcome measures.

Multilevel modelling has provided a framework within which to analyse data obtained from studies with complex sample designs. Structural equation modelling has provided a framework within which to analyse truly multivariate problems and in which account can be taken of measurement error. In practice many studies have required analysis that recognizes both problems, but until recently researchers have been forced to ignore or to treat very pragmatically one or the other. Such compromise should no longer be necessary.

We have attempted to emphasize the commonalities between multilevel and latent variable modelling, but we have not exhausted this theme. For example, factor score estimation and random effects estimation share a common basis, with the usual estimators exhibiting 'shrinkage' in both cases. This commonality also extends to areas of relative weakness, models diagnostics being a case in point.

The fusion of random effects and latent variable modelling has resulted in other novel possibilities. Modelling with non-normal random effects has been explored both within Bayesian Monte-Carlo Markov Chain Estimation and in ML estimation, where both parametric and non-parametric (e.g. NPML) estimators have been considered.[65,66] Rabe-Hesketh and Pickles (submitted) have shown that the properties of NPML estimators for random effects models also apply to circumstances where that random effect is a latent variable, that is to models involving factor loadings. Since latent class analysis can be considered as a simplified or approximate NPML estimator then multilevel latent class models can be investigated within this framework.

Acknowledgements

The work of SY and AP was financially supported by grant H519255050 from the UK ESRC. We would like to thank Anders Skrondal for his helpful comments.

References

1 Goldstein H, Rasbash J, Plewis I *et al*. *A user's guide to MLwiN*. London: Multilevel Models Project, Institute of Education, University of London, 1998.
2 Bryk AS, Raudenbush SW, Congdon R. *HLM. Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. Chicago, IL: Scientific Software International, 1996.
3 Longford NT. *VARCL: Software for variance component analysis of data with nested random effects (maximum likelihood)*. Groningen: ProGamma, 1993.
4 Bryk AS, Raudenbush SW. *Hierarchical linear models, applications and data analysis methods*. Newsbury Park, CA: Sage, 1992.
5 Longford NT. *Random coeffecient models*. Oxford: Oxford University Press, 1993.
6 Kreft I, De Leeuw J. *Introducing multilevel modeling*. London: Sage, 1998.
7 Snijders TAB, Bosker RJ. *Multilevel analysis*. London: Sage, 1999.

8 Hox JJ. *Applied multilevel analysis*. Amsterdam: TT-Publications, 1995.
9 Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**: 9–25.
10 Ezzet F, Whitehead J. A random effects model for ordinal responses from a cross-over trial. *Statistics in Medicine* 1991; **10**: 901–907.
11 Hedeker D, Gibbons RD. MIXOR: a computer program for mixed effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine* 1996; **49**: 157–76.
12 Rabe-Hesketh S, Pickles A, Taylor C. sg129: generalized linear latent and mixed models. *Stata Technical Bulletin* 2000; **53**: 47–57.
13 Liu Q, Pierce DA. A note on gauss-hermite quadrature. *Biometrika* 1994; **81**: 624–29.
14 Zeger SL, Karim MR. Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Society* 1991; **86**: 79–102.
15 Keane MP. A computationally practical simulation estimator for panel data. *Econometrica* 1994; **62**: 95–116.
16 McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 1997; **92**: 162–70.
17 Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**: 1049–60.
18 Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford University Press, 1994.
19 McCullagh P. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B* 1980; **42**: 109–42.
20 Bennett S. Analysis of survival data by the proportional odds model. *Statistics in Medicine* 1983; **2**: 273–77.
21 Hedeker D, Siddiqui O, Hu FB. Random-effects regression analysis of correlated grouped-time survival data. *Statistical Methods in Medical Research* 2000; **9**: 161–79.
22 Agresti A. *Analysis of ordinal categorical data*. New York: John Wiley, 1984.
23 Jenkins SP. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* 1995; **57**: 129–37.
24 Singer JD, Willett JB. It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Education Statistics* 1993; **18**: 155–95.
25 Thompson WA. On the treatment of grouped observations in life studies. *Biometrics* 1977; **33**: 463–70.
26 Collett D. *Modelling survival data in medical research*. London: Chapman & Hall, 1994.
27 Hosmer DW, Lemeshow S. *Applied survival analysis: regression modeling of time to event data*. New York: John Wiley, 1999.
28 Scheike TH, Jensen TK. A discrete survival model with random effects: an application to time to pregnancy. *Biometrics* 1997; **53**: 318–29.
29 Muthén LK, Muthén BO. *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén, 1998.
30 Neale MC. *Mx: statistical modeling, 4th edn*. Richmond, VA: Virginia Commonwealth University, Medical College of Virginia, 1997.
31 Lader D, Matheson J. *Smoking among secondary school children in 1990*. London: HMSO, 1991.
32 Pickles A, Pickering K, Taylor C, Sutton C, Yang S. Multilevel risk models for retrospective age-of-onset data: school children's first cigarette. *Journal of Adolescent Research* 2001; **16**: 188–204.
33 Rabe-Hesketh S, Pickles A. Generalized linear latent and mixed models. In: Friedl H, Berghold A, Kauermann G, eds. *Proceedings of the 14th International Workshop on Statistical Modeling*. Graz, Austria, 1999: 332–39.
34 Sudman S, Bradburn NM. Effects of time and memory factor on response in surveys. *Journal of the American Statistical Association* 1973; **63**: 805–15.
35 Hobcraft J, Murphy M. Demographic event history analysis: a selective review. In: Crouchley R, ed. *Longitudinal data analysis*. Aldershot: Avebury, 1987.
36 Pickles A, Pickering K, Taylor C. Reconciling recalled dates of developmental milestones, events and transitions: a mixed generalized linear model with random mean and variance functions. *Statistics in Medicine* 1995; **14**: 1447–61.
37 Yang S, Pickles A, Taylor C. Multilevel structural equation models for complex data. In: *Proceedings of ASC'99 – Leading Survey and Statistical Computing into the New Millennium*. Edinburgh, 1999: 22–24.
38 Yang S, Pickles A, Taylor C. Multilevel latent variable model for analysing two-phase survey data. In: Friedl H, Berghold A, Kauermann G, eds. *Proceedings fo the 14th International Workshop*

on Statistical Modeling. Graz, Austria, 1999: 402–408.
39 Goldstein H. Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika* 1986; **73**: 43–56.
40 Goldstein H. Restricted unbiased iterative generalised least squares estimation. *Biometrika* 1989; **76**: 622–23.
41 Rabe-Hesketh S, Pickles A, Skrondal A. Gllamm: a general class of multilevel models and a stata program. *Multilevel Modelling Newsletter* 2001; **13**: 17–23.
42 Dunn G, Everitt B, Pickles A. *Modelling covariances and latent variables using EQS*. London: Chapman & Hall, 1993.
43 Bollen KA. *Structural equations with latent variables*. New York: John Wiley, 1989.
44 Bentler PM. *EQS structural equation program manual*. Encino, CA: Multivariate Software, 1995.
45 Arbruckle J. *Amos' users' guide*. Chicago, IL: SmallWaters Corporation, 1997.
46 Jöreskog K, Sörbom D. *LISREL 8: structural equation modeling with the SIMPLIS command language*. Hove: Scientific International, 1996.
47 Muthén BO. A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika* 1984; **49**: 115–32.
48 Little TD, Schnabel KU, Baumert J. *Modeling longitudinal and multilevel data*. Mahwah, NJ: Lawrence Erlbaum, 2000.
49 Muthén BO. On structural equation modeling with data that are not missing completely at random. *Psychometrika* 1987; **52**: 431–62.
50 Longford NT, Muthén BO. Factor analysis for clustered observations. *Psychometrika* 1992; **57**: 581–97.
51 Poon WY, Lee SY. Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statistics and Probability Letters* 1992; **14**: 25–30.
52 Linda NY, Lee YSY, Poon WY. Covariance structure analysis with three level data. *Computational Statistics and Data Analysis* 1993; **15**: 159–78.
53 Muthén BO, Satorra A. Complex sample data in structural equation modeling. In: Marsden P, ed. *Sociological methodology*. Cambridge, MA: Blackwell, 1995: 267–316.
54 Goldstein H, McDonald RP. A general model for the analysis of multilevel data. *Psychometrika* 1988; **53**: 455–67.
55 McDonald RP, Goldstein H. Balanced and unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology* 1989; **42**: 215–32.
56 Muthén BO. Latent variable modeling of longitudinal and multilevel data. In: Raftery AE, ed. *Sociological Methodology*. Cambridge, MA: Blackwell, 1997: 453–80.
57 Muthén BO. Latent variable modeling in heterogeneous populations. *Psychometrika* 1989; **54**: 557–85.
58 Muthén BO. Multilevel covariance structure analysis. *Sociological Methods and Research 1994;* **22***: 376–98.*
59 Raudenbush SW. Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology* 1995; **48**: 359–70.
60 Lee SY, Tsang SY. Constrained maximum likelihood estimation of two-level covariance structure model via EM type algorithms. *Psychometrika* 1999; 64: 435–50.
61 Goldberg D. Why GHQ thresholds vary from one place to another. *Psychological Medicine* 1994; **28**: 915–21.
62 Ustun B, Sartorius N. *Mental illness in general health care: an international study*. Chichester: John Wiley, 1995.
63 Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–92.
64 Neale APM, Simonoff E, Rutter M *et al*. A simple method for censored age of onset data subject to recall bias: mother's reports of age of puberty in male twins. *Behavior Genetics* 1994; **24**: 457–68.
65 Simar L. Maximum likelihood estimation of a compound Poission process. *The Annals of Statistics* 1976; **4**: 1200–209.
66 Aitkin M. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 1999; **55**: 117–28.

Appendix: model setup of the direct ML approach

The multilevel latent growth curve model for the GHQ data in (25) can be written in the form of equation (19), as

$$\mathbf{y}_{ij} = \mathbf{X}\boldsymbol{\nu} + \mathbf{H}_2\mathbf{u}_{ij}^{(2)} + \mathbf{H}_3\mathbf{u}_{ij}^{(3)} + \mathbf{H}_4\mathbf{u}_{i}^{(4)} + \boldsymbol{\epsilon}_{ij} \tag{26}$$

For subjects who have complete data, i.e. those who have all of the nine measurements, this becomes

$$\begin{bmatrix} y_{ij11} \\ y_{ij12} \\ y_{ij21} \\ y_{ij22} \\ y_{ij23} \\ y_{ij31} \\ y_{ij33} \\ y_{ij41} \\ y_{ij43} \end{bmatrix} = \mathbf{I}_9\boldsymbol{\nu} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} u_{0ij1}^{(2)} \\ u_{0ij2}^{(2)} \\ u_{0ij3}^{(2)} \\ u_{0ij4}^{(2)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_{0ij}^{(3)} \\ u_{1ij}^{(3)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_{0i}^{(4)} \\ u_{1i}^{(4)} \end{bmatrix} + \begin{bmatrix} \epsilon_{ij11} \\ \epsilon_{ij12} \\ \epsilon_{ij21} \\ \epsilon_{ij22} \\ \epsilon_{ij23} \\ \epsilon_{ij31} \\ \epsilon_{ij33} \\ \epsilon_{ij41} \\ \epsilon_{ij43} \end{bmatrix} \tag{27}$$

where $\mathbf{I}_9$ is an 9×9 identity matrix and the elements of $\boldsymbol{\nu}$ are the intercepts ν_{rs}. Assume that $\boldsymbol{\Sigma}, \Psi, \boldsymbol{\Sigma}^{(3)}$ and $\boldsymbol{\Sigma}^{(4)}$ are variance covariance matrices of $\boldsymbol{\epsilon}_{ij}, \mathbf{u}_{ij}^{(2)}, \mathbf{u}_{ij}^{(3)}$ and $\mathbf{u}_{i}^{(4)}$, respectively. As described in Section 3.3, it is assumed that ϵ_{ijrs} are uncorrelated, and the variance of ϵ_{ijrs} equals $\sigma_s^2 (s = 1, 2, 3)$. $u_{0ijr}^{(2)}$ are mutually independent with a common variance, ψ. The format of $\boldsymbol{\Sigma}, \Psi, \boldsymbol{\Sigma}^{(3)}$ and $\boldsymbol{\Sigma}^{(4)}$ are as follows:

$$\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_1^2, \sigma_3^2, \sigma_1^2, \sigma_3^2) \tag{28}$$

$$\Psi = \text{diag}(\psi, \psi, \psi, \psi) \tag{29}$$

$$\boldsymbol{\Sigma}^{(3)} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{(3)} & \boldsymbol{\Sigma}_{12}^{(3)} \\ \boldsymbol{\Sigma}_{12}^{(3)} & \boldsymbol{\Sigma}_{22}^{(3)} \end{bmatrix} \tag{30}$$

and

$$\boldsymbol{\Sigma}^{(4)} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{(4)} & \boldsymbol{\Sigma}_{12}^{(4)} \\ \boldsymbol{\Sigma}_{12}^{(4)} & \boldsymbol{\Sigma}_{22}^{(4)} \end{bmatrix} \tag{31}$$

The unknown parameters of this model are ν, λ_2, λ_3, and

$$\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \psi, \boldsymbol{\Sigma}_{11}^{(3)}, \boldsymbol{\Sigma}_{12}^{(3)}, \boldsymbol{\Sigma}_{22}^{(3)}, \boldsymbol{\Sigma}_{11}^{(4)}, \boldsymbol{\Sigma}_{12}^{(4)}, \boldsymbol{\Sigma}_{22}^{(4)})$$

For cases with an incomplete set of response measures the corresponding rows in the above equation are extracted when the likelihood is computed. For example, only the first two rows in the equation contribute to the likelihood for cases with the first-phase screening variables y_{11} and y_{12} only. The estimation procedure first estimates ν using least squares, then the likelihood is maximized with respect to the factor loadings λ_2 and λ_3, and finally the parameters in the covariance matrices, $\boldsymbol{\sigma}$.