Longitudinal and Life Course Studies 2010 Volume 1 Issue 2 pp 170 - 185

Statistical modelling of repeated measurement data

Harvey Goldstein
University of Bristol and London School of Hygiene and Tropical Medicine
h.goldstein@bristol.ac.uk

Bianca De Stavola
London School of Hygiene and Tropical Medicine

Abstract

This tutorial describes ways of modelling repeated measurements taken on a sample of individuals. It gives a brief historical introduction and then describes how a 2-level formulation provides a flexible and straightforward approach.

Keywords

Growth curves, multilevel model, multivariate analysis of variance, repeated measures models, smoothing models

1. Introduction

The fitting of statistical models to sequential or repeated measurements over time on the same individuals, has a long history. Early interest centred on characterising the growth period of individuals, children and animals, and a discussion can be found in Tanner (1979). These were attempts to fit smoothly varying curves to the growth period, and by the 1930s had spawned a large literature that included complex non-linear models with several parameters that many investigators claimed could be associated with 'biologically meaningful' characteristics (see Goldstein, 1979 for a discussion). These procedures carried out separate fitting for each individual's set of measurements. Subsequent developments generalised these models to consider samples of individuals where terms or 'parameters' are included in order to account for the between-individual variability in growth patterns. In the case of non-linear models the work of Bock (1989) is notable.

The main feature that distinguishes approaches to the fitting of models to repeated measurements is the actual structure of the repeated measures. If observations are planned to occur at the same occasions for all individuals the data are said to be *balanced*. By contrast, if observations occur at irregular time points they are *unbalanced*. The first setting gives rise not only to equal numbers of observations on each individual, but also to a choice of modelling approaches because the data can be viewed as arising either from correlated observations of a single response or 'dependent' variable, or from multivariate outcomes (each outcome associated with one observation time). With unbalanced data only the first approach is generally possible.

In this tutorial we shall give an overview of the more commonly used methods to model repeated measurement data, distinguishing between these two main settings. We shall also touch upon the issue of missing (incomplete) data. We illustrate the various approaches by fitting alternative models to some growth data collected in the `Oxford Boys Study' [Goldstein et al., 1994].

2. Analyses based on balanced data

We set out some basic statistical models for repeated measurement data starting with the balanced case. Our response variable is denoted by $\boldsymbol{Y}$ and the *i*-th response for the *j*-th individual in a sample is denoted by y_{ij}. Suppose that there are *p*

measurement occasions at times t_1, t_2, …, t_p, for each individual *j, j=1, …,N,* and we have complete balanced data. To fix ideas with an example we shall consider the case of 3 occasions, with obvious generalisations, so that the data will look like those in Table 1.

Table 1. Balanced data example with p=3 repeated observations taken at fixed occasion times

Time / Individual	Occasion 1 t_1	Occasion 2 t_2	Occasion 3 t_3
1	y_{11}	y_{21}	y_{31}
2	y_{12}	y_{22}	y_{32}
….	….	….	….
j	y_{1j}	y_{2j}	y_{3j}
….	….	….	….
N	y_{1n}	y_{2n}	y_{3n}

Data missingness, i.e. settings where some of the cells in this table are empty, will be discussed later.

2.1 Multivariate model formulation

Multivariate analysis of variance models for repeated measures, especially those associated with the work of Rao (1965), are summarised in Grizzle and Allen (1969). In brief, a general model is formulated as follows. We write

$$
\begin{aligned}
y_{1j} &= \beta_0 + \beta_1 t_1 + e_{1j} \\
y_{2j} &= \beta_0 + \beta_1 t_2 + e_{2j} \\
y_{3j} &= \beta_0 + \beta_1 t_3 + e_{3j} \\
\begin{pmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \end{pmatrix} &\sim MVN(0, \Omega), \quad \Omega = \begin{pmatrix} \sigma^2_{e1} & & \\ \sigma_{e12} & \sigma^2_{e2} & \\ \sigma_{e13} & \sigma_{e23} & \sigma^2_{e3} \end{pmatrix}
\end{aligned}
\qquad (1)
$$

In other words the three measurements on *y* are assumed to have a multivariate normal (*MVN*) distribution, in this case a 3-variate distribution. Model (1) is therefore simply a multivariate analysis of variance with a single covariate *t* and thus assumes linear growth. Clearly this can be generalised to include higher order polynomial terms and other covariates measured on individuals such as gender or birthweight and even time varying covariates. The innovations introduced in the 1960s involved considering particular structures for Ω. Thus, Rao (1965) considers a structure where each individual has his/her specific coefficients and these coefficients vary across

individuals, with independent residuals having a common variance. This gives the structure

$$\Omega = T\Omega_u T^T + \sigma_e^2 I \qquad (2)$$

where ***I*** is the identity matrix and

$$T = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \qquad \Omega_u = \begin{pmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{pmatrix} \qquad (3)$$

This specification of the variance matrix Ω leads to separating two components of variation: the first, $T\Omega_u T^T$ captures how growth trajectories vary across individuals via the definition of Ω_u, and the second captures the within individual 'noise' via the residual variance σ^2_e. Note also that the first component varies quadratically with time. We shall have more to say about this structure when we discuss the multilevel model formulation.

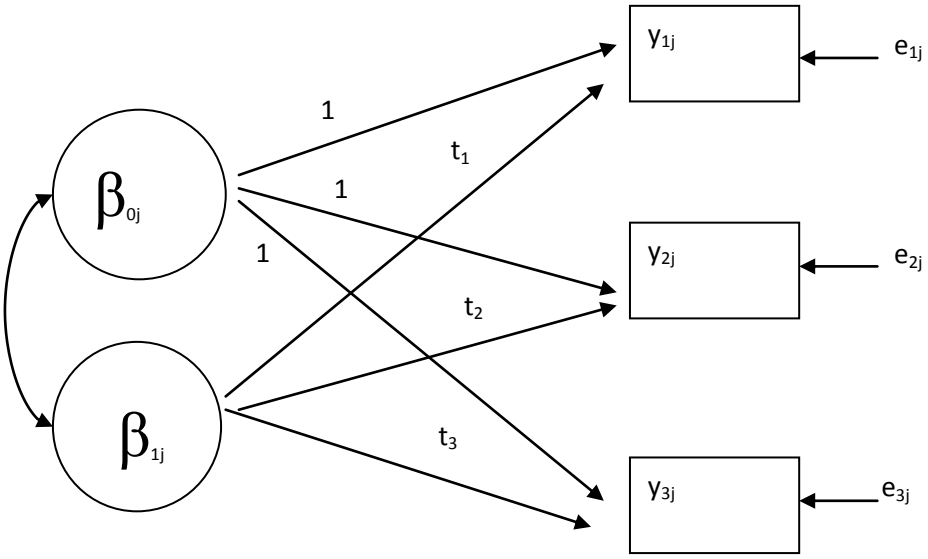
2.2 Latent variables formulation

Model (1) can also be written as

$$\begin{pmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \begin{pmatrix} \beta_{1j} \\ \beta_{2j} \end{pmatrix} + \begin{pmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \end{pmatrix} = \mathrm{T}^{\mathrm{T}}\ \beta + \mathrm{e}, \quad \begin{pmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \end{pmatrix} \sim MVN\ (T^T\beta, \Omega\ ), \quad \Omega = T\Omega_u T^T + \Omega_e \qquad (4)$$

We have now introduced the subject specific intercept β_{1j} and slope β_{2j} which are treated as latent variables with known regression coefficients (as defined by the matrix T) and variance covariance matrix Ω_u while the errors e_{ij} have variance covariance matrix Ω_e. A flexibility allowed by this model is that Ω_e is not forced to have identical variances at each occasion, i.e. we can allow $\Omega_e \neq \sigma^2_e \boldsymbol{I}$. The model is often represented using a path diagram as shown in Figure 1, which applies to the example with repeated observations at three fixed times. As usual in this field, latent variables are represented by circles and observed (`manifest') variables by rectangles. Arrows indicate the direction of assumed association. Note that the regression coefficients (in this literature termed `*factor loadings*') are fixed in the matrix T. The covariance between the two latent variables (represented by a double arrow joining the two variables) corresponds to σ_{u01} in the covariance matrix of equation (3).

Figure 1 - **Path analytical representation for the latent variables formulation**



2.3 Analyses based on unbalanced data

As before, denote the *i*-th measurement for the *j*-th individual by y_{ij}. Suppose that there are p_j measurement occasions for the *j*-th individual. To fix ideas with an example we shall consider the case depicted in Table 2. Here individuals are observed at different time points and have different total numbers of observations. Such a data structure could be viewed as a balanced one which is affected by missingness, i.e. where potentially everybody is observed at all listed time points t_{ij}, for all individuals *j*=1, …,*N*. Note, however, that this is still not fully general, since we are assuming a fixed number of discrete time points. In the next section, and in our example, we shall allow the time points (or ages) to occur anywhere.

.Table 2. Unbalanced data example with varying number of observations per individuals pj out of a maximum of 5 possible time points

		Occasion					
Individual	**Variable**	**1**	**2**	**3**	**4**	**5**	**pj**
1	**Time**	t_{11}	t_{21}	-	-	-	-
	Y	y_{11}	y_{21}	-	-	-	**2**
2	**Time**	t_{12}	t_{22}	-	t_{42}	-	-
	Y	y_{12}	y_{22}	-	y_{42}	-	**3**
…	**…**	**…**	**…**	**…**	**…**	**…**	**…**
	…	**…**	**…**	**…**	**…**	**…**	**…**
j	**Time**	t_{1j}	t_{2j}	t_{3j}	-	t_{5j}	
	Y	y_{1j}	y_{2j}	y_{3j}	-	y_{5j}	**4**
…	**…**	**….**	**….**	**….**	**…**		

For such unbalanced designs, assuming that the data are 'missing' at random, modifications to the standard multivariate or latent variable analyses are available (Muthen, 1997)

3. The multilevel model for repeated measures

The seminal paper that introduces the modern approach to fitting repeated measures data is that of Laird and Ware (1982). In essence their model - often referred to as a linear mixed or random coefficient model - can be written, analogously to (1) – (3) as

$$
\begin{aligned}
y_{1j} &= \beta_{0j} + \beta_{1j}t_{1j} + e_{1j} \\
y_{2j} &= \beta_{0j} + \beta_{1j}t_{2j} + e_{2j} \\
y_{3j} &= \beta_{0j} + \beta_{1j}t_{3j} + e_{3j} \\
\beta_{0j} &= \beta_0 + u_{0j} \\
\beta_{1j} &= \beta_1 + u_{1j} \\
\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} &\sim MVN(0,\Omega_u), \quad \Omega_u = \begin{pmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{pmatrix} \\
\begin{pmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \end{pmatrix} &\sim MVN(0,\Omega_e), \quad \Omega_e = \sigma^2_e I = \begin{pmatrix} \sigma^2_e & & \\ 0 & \sigma^2_e & \\ 0 & 0 & \sigma^2_e \end{pmatrix}
\end{aligned}
\qquad (5)
$$

The essential difference between the earlier specification and (5) is that in (5) the time points are allowed to be quite general, that is to differ across individuals so that individuals need not have the same number or spacing of time points. This flexibility is available because (5) is essentially a univariate model where the response is directly modelled as a function of time. It is in fact a 2-level model and thus a special case of a general multilevel model (see Goldstein, 2003). As before extensions include higher order polynomial terms for *t* and other explanatory variables.

3.1 Semi-parametric modelling

Alternative approaches to parametric modelling of repeated measurement data, using the multilevel formulation, involve fitting smoothing or regression splines. Often, the results generally are best presented graphically, although features of the fitted growth curves at chosen ages can be derived numerically. We now show how these can be fitted, concentrating on the regression spline model.

3.2 Smoothing splines

Smoothing splines are non-parametric functions of the outcome variable, *Y*, on time, T, that are selected according to a 'roughness penalty'. As before, let y_{ij} be the *i*-th measurement for the *j*-th individual taken at time t_{ij}. A simple example of a smoothing spline would be a running mean based on a moving window (of size *h*) which is placed symmetrically around each data point, t_{ij}. The estimate of the outcome specific to the point at the centre of the window is then calculated as the mean of the outcome values belonging to that window or as the prediction from a least squares model fitted to those data points.

More general approaches produce predictions based on weighted functions, called *kernels*, across the whole range of data, with points closer to the centre of the kernel given the greatest influence. Changes in the chosen kernel function will lead to changes in the fit to the observed data. However, any improvement in fit may lead to a very jagged shape so that roughness penalties that provide a compromise between fit and smoothness, are used to guide the selection of kernel function. This leads to penalized least squares estimators and automatic selection procedures such as cross-validation [Green and Silverman, 1993], and these procedures can be fitted within the framework of a multilevel or random coefficient model; see for example Wang (1998). In the next section we describe a flexible system that is easily embedded within a multilevel framework.

3.3 Regression splines

The simplest example of a regression spline is a piecewise linear model where simple linear regression models are fitted within consecutive intervals, defined by so-called *knots*, with the lines joining at the knots. More general models include quadratic and cubic splines, with the latter being most commonly used [Pan and Goldstein, 1998]. In such cases the joins can be made 'smooth' in the sense defined below. The main problem with this approach is the selection of the number and location of the knots to be used.

A simple piecewise quadratic version of this model, with a single knot at time t_0, can be written as follows:

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 (t_i - t_0)_+^2 + e_i$$
$$e_i \sim N(0, \sigma_e^2) \qquad (6)$$
$$(t_i - t_0)_+ = \begin{cases} (t_i - t_0) & t_i \geq t_0 \\ 0 & t_i < t_0 \end{cases}$$

where the term $(t_i\text{-}t_0)_+$ is 'grafted' onto the existing polynomial. In (6) for simplicity we have dropped the suffix *j*. This function has continuous first derivatives and therefore will be smooth at the location of the knot. Thus, at time t_0 the predicted value is $\beta_0 + \beta_1 t_0 + \beta_2 t_0^2$ and the first derivative, the rate of change with time, is $t_1 + 2\beta_2 t_0$.

Thus, at a very small later time $x = t_0 + \Delta$ the predicted value is $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 \Delta^2$, that is a point very slightly perturbed away from the quadratic curve defined at time t_0 .

A direct generalization of this model to a multilevel structure with intercept and slope terms varying across individuals can be written as:

$$
\begin{aligned}
&y_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + \beta_2 t_{ij}^2 + \beta_3 (t_{ij} - t_0)_+^2 + e_{ij} \\
&\beta_{0j} = \beta_0 + u_{0j} \\
&\beta_{1j} = \beta_1 + u_{1j} \\
&\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim MVN(0, \Omega_u), \quad \Omega_u = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix} \\
&e_{ij} \sim N(0, \sigma_e^2) \\
&(t_{ij} - t_0)_+ = \begin{cases} (t_{ij} - t_0) & t_{ij} \geq t_0 \\ 0 & t_{ij} < t_0 \end{cases}
\end{aligned}
\qquad (7)
$$

This is just a 2-level model with an additional quadratic term smoothly 'grafted' onto the average quadratic relationship at time t_0, and with each individual having a subject-specific intercept β_0 and subject-specific slope for time β_1. For simplicity here the quadratic coefficients β_2 and β_3 are assumed not to vary across individuals, but either or both could be made random.

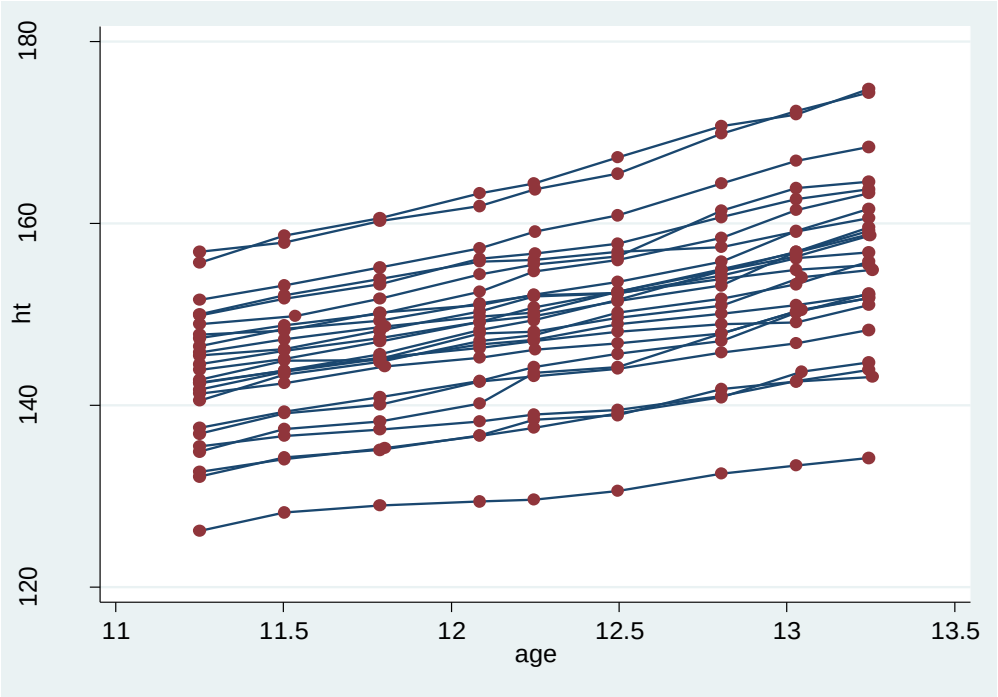
4. Estimation and software

For continuous normal response measurements these models are readily fitted using maximum likelihood (ML). With balanced data, this gives unbiased estimates of β although, especially where there are relatively small numbers of individuals, somewhat biased estimates of the parameters in Ω. In such cases we can use restricted maximum likelihood estimation (REML) which gives unbiased estimates of all the parameters (see for example, Rabe-Hesketh and Skrondal, 2008). REML estimation in its simplest form is the familiar use of the divisor *n-1* for a variance rather than *n* which is the maximum likelihood estimator. Multivariate models can be fitted using SAS, Stata, SPSS and specialised software such as Mplus and MLwiN. All but MPlus allow a general 2-level model to be fitted with both ML and REML estimation.

5. Example: The Oxford boys data

The dataset consists of repeated measures of the height of 26 boys taken on 9 occasions each over a 2 year spell from a residential school in Oxfordshire, England. (Available as an EXCEL spreadsheet through the 'Reading Tools' link on the right of the screen when this paper is viewed). The boys were just over 11 years of age at entry. The age scale in the following models is centred at 12.25 years. Figure 2 shows their observed growth profiles and highlights the regularity of their observed time points, with only a few exceptions. There is clearly a ranking in height at entry that is generally maintained over the full observation period.

Figure 2. Observed growth profiles



The boys seem to increase their rate of growth with time, from around a mean of 6 cm/year between about age 11 and 12 years to nearly 9 cm/years shown in Table 3. The standard deviation of growth also seems to increase with time.

Table 3 – Mean and standard deviation (SD) of the yearly rate of increase (*cm/year*) between successive visits by age interval – Oxford Boys Study

Age interval (years)	*Mean*	*SD*
11 ¼ to 11 ½	6.53	2.53
11 ½ to 11 ¾	5.05	1.91
11 ¾ to 12	6.21	1.28
12 to 12 ¼	6.77	4.50
12 ¼ to 12 ½	5.31	2.75
12 ½ to 12 ¾	7.26	3.58
12 ¾ to 13	8.95	4.01
13 to 13 ¼	7.23	3.22

The repeated observations are effectively balanced, as every boy was observed 9 times, roughly every 3 months. Note the high correlation among these repeated observations as well as its strengthening when they are closest in time.

Table 4. Observed correlations between height measured at 9 visits - Oxford Boys Study.

		Age at visit (years)								
		11 ¼	11 ½	11 ¾	12	12 ¼	12 ½	12 ¾	13	13 ¼
Age at visit (years)	11 ¼	1.000								
	11 ½	0.996	1.000							
	11 ¾	0.997	0.999	1.000						
	12	0.992	0.996	0.997	1.000					
	12 ¼	0.987	0.991	0.992	0.996	1.000				
	12 ½	0.983	0.989	0.991	0.995	0.997	1.000			
	12 ¾	0.970	0.974	0.977	0.983	0.991	0.993	1.000		
	13	0.962	0.963	0.968	0.975	0.984	0.988	0.995	1.000	
	13 ¼	0.957	0.959	0.964	0.970	0.979	0.986	0.992	0.998	1.000

We can fit model (1) - or equivalently model (4) – to these data by treating the 9 observations as if they were taken at exactly the same set of 3-monthly intervals (which is nearly correct- see Figure 2). We do this by specifying the (9 x 2) matrix T in equation (4) as a column of 1s and a column where the common observation times measured in years t_i , i=1,..9, are: (-1.0, -0.75, -0.50, -0.25, 0.0, 0.25, 0.50, 0.75, 1)

The model is centred at age 12 ¼ years. This choice of values will lead to the intercept term referring to the mid observation. The results are shown in Table 5 where we fit both a linear model and one that includes a quadratic term in the fixed part of the model that describes average growth. With the first specification we estimate that the random intercept β_0 has mean 149.5 cm and variance 63.2 cm^2 while the linear slope β_1 has mean 6.5 cm/year and variance 2.7 $(cm/year)^2$. The random intercept and random slope have covariance 8.41 and a corresponding correlation of 0.64. Comparisons of log likelihood values indicate that specification 2 with a quadratic term gives a better fit to the data (deviance=21.7 to be judged according to a chi-squared distribution with 1 degree of freedom, p<0.001).

Table 5 – Alternative specifications of model (4) with 2 latent variables β_0 and β_1: as a linear function of age (specification 1) and as a quadratic function of age (specification 2): ML estimation. Estimates and standard errors (SE).

	Specification 1		*Specification 2*	
	Coef	SE	Coef	SE
Latent variable:				
β_0 (intercept)				
Mean	149.50	1.56	149.30	1.56
Variance	63.16	17.53	63.16	17.53
β_1 (linear effect of age (*yrs, centred*))				
Mean	6.54	0.33	6.54	0.33
Variance	2.72	0.78	2.74	0.79
Fixed effect:				
Age^2 (yrs, centred)	-	-	0.53	0.11
$Cov(\beta_0,\beta_1)$	8.41	3.10	8.41	3.10
Residual variance(s) level 1	0.42	0.05	0.38	0.04
-2 Log likelihood	721.28		699.55	

Exactly the same results as those of table 5 can be obtained using a multilevel approach with the actual times of observations replaced by their planned times, ie 11 ¼, 11 ½ , etc., centred around age 12 ¼ years. The first two columns of table 6 report results in the format usually adopted with multilevel models. Here we also compare the results obtained using ML and REML estimation. Because the sample size is relatively small (26 boys measured 9 times) the estimates of the random part of the model do differ somewhat.

We now fit a multilevel model where we use the exact age at the measurement occasions (third set of columns in Table 6). There are some small differences in estimated values from those obtained using the planned ages due to the fact that there is only slight variation from the target ages of measurement.

Table 6 - Alternative specifications for model (5)

	Multilevel growth model with time defined by:					
	Planned age at visit				Observed age at visit	
	ML		REML		REML	
	Coef	SE	Coef	SE	Coef	SE
Fixed part						
Intercept	149.52	1.56	149.52	1.59	149.37	1.59
Age (yrs, centred)	6.54	0.33	6.54	0.34	6.53	0.34
Random part						
$\sigma^2_{Intercept}$	63.15	17.53	65.68	1.59	65.30	18.49
σ^2_{Age}	2.72	0.79	2.84	0.84	2.83	0.83
$\sigma_{Intercept-Age}$	8.41	3.10	8.75	3.29	8.71	3.27
Residual variance	0.42	0.05	0.42	0.05	0.44	0.05
-2 Log likelihood	721.28		719.40		724.08	

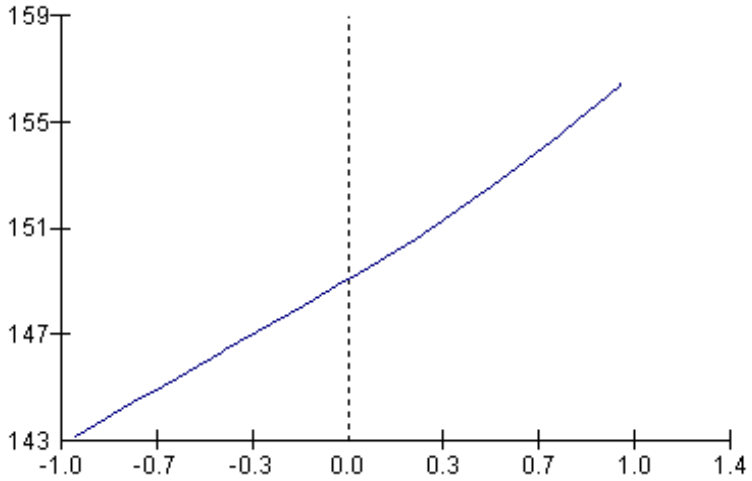
As before, adding a quadratic term to the last model we fitted, we find that it is significant (see first set of columns in Table 7).

We can compare the model with a quadratic term with one obtained using a multilevel quadratic regression spline with a single knot at the centred age of zero (i.e. age 12 ¼ on the original scale), which is an example of model (7). We fit two specifications of this model: the first has both the standard quadratic term in age and the additional term identifying the local departure from the quadratic function after the knot at 0. The second removes the original quadratic term as it was found not to be significant. Figure 3 shows the predicted average growth curve for this final model. The quadratic component only comes in at the mean age of 12 ¼ years: before that the growth is effectively linear. This illustrates the flexibility that we can introduce within the class of models that are based upon polynomials.

Table 7- Alternative specifications of model (5) and model (7) – time defined as the observed age at visit (in years, centred) and REML estimation

	Multilevel model		**Regression spline model**			
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Fixed part						
Intercept	149.06	1.56	149.06	1.59	149.06	1.59
Age (yrs, centred)	6.52	0.33	5.93	0.42	5.89	0.35
Age^2 (yrs, centred)	0.74	0.10	0.05	0.31	-	
$(Age)_+^{\zeta}$	-	-	1.40	0.59	1.49	0.19
Random part						
$\sigma^2_{Intercept}$	62.81	17.43	65.32	18.49	65.32	18.49
σ^2_{Age}	2.74	0.78	2.85	0.83	2.85	0.83
$\sigma_{Intercept-Age}$	8.38	3.09	8.72	3.28	8.72	3.28
Residual variance	0.44	0.05	0.33	0.04	0.33	0.03
-2 Log likelihood	780.20		674.74		674.26	

Figure 3. Predicted average growth curve for quadratic regression spline model with linear growth before 12 ¼ years.



6. Checking assumptions

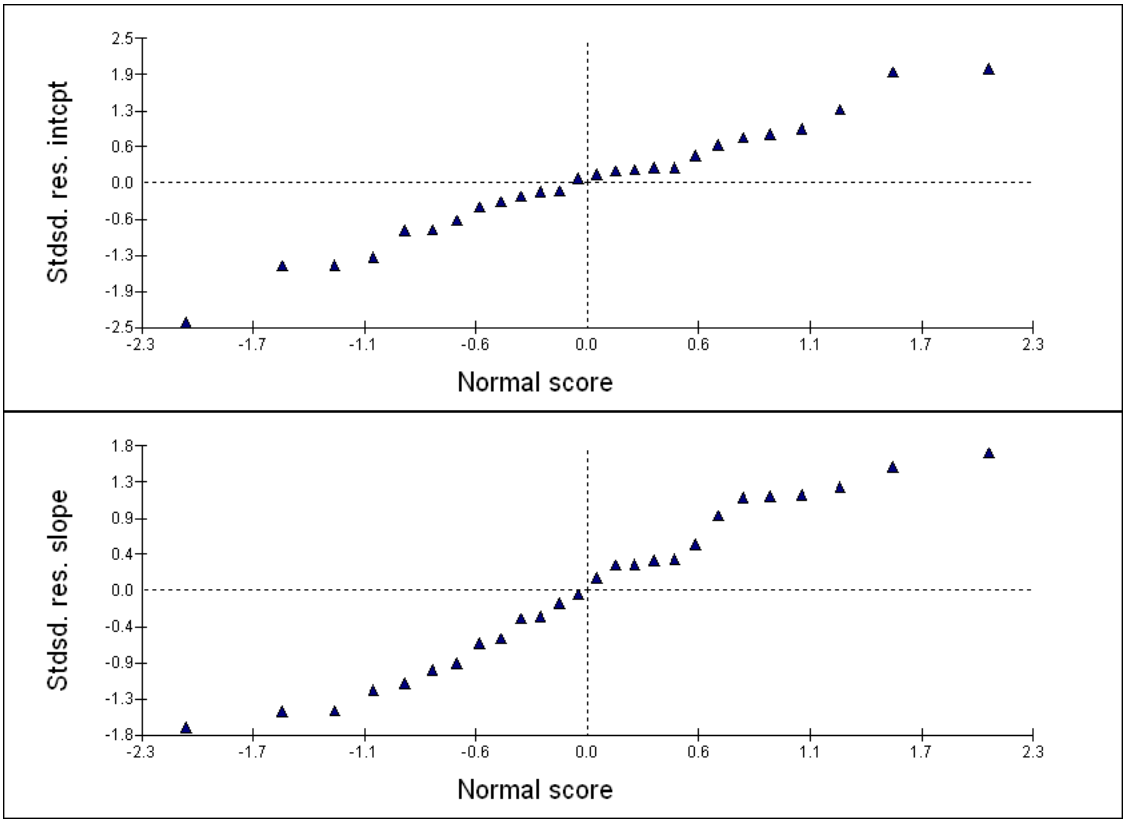
Several diagnostic procedures are available for multilevel models to check the various assumptions made, especially that of normality (See Goldstein, 2003 Chapter 2 for more details). To illustrate, Figure 4 shows normal quantile plots for the level 2 (individual) and level 1 (occasion) estimated standardised residuals. If the assumption of normality for the random effects (residuals) is correct these plots should be approximately linear. At level 2 there is some evidence of departure from normality for the slope estimates, but little evidence at level 1. In particular there are no estimates that appear to be real outliers.

In fact, we can go on and allow the coefficient of the grafted quadratic term to vary randomly over individuals and this does improve the fit. We omit the details but it produces more linear plots at level 2 as shown in Figure 5.

A further assumption that we have made is that the residuals, especially those at level 1 are independent. In some cases, for example when measurements are taken very close together, this may not be the case and we may need to fit, say, a model with autocorrelated level 1 residuals. For a discussion of this case with an example see Goldstein et al., (1994).

Figure 4. Standardised residuals for model in column 3 in Table 7.

Level 2



Level 1

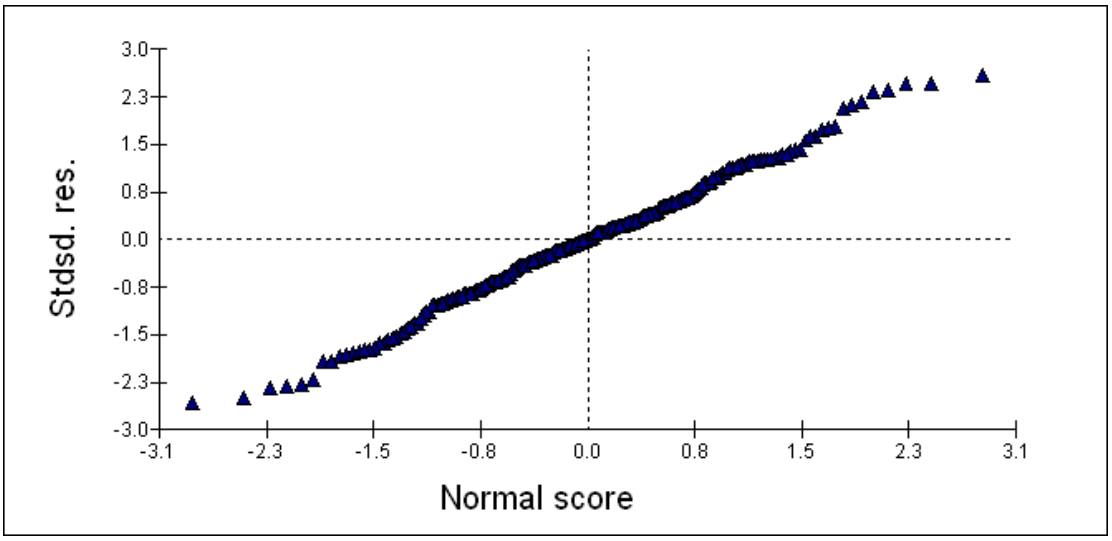
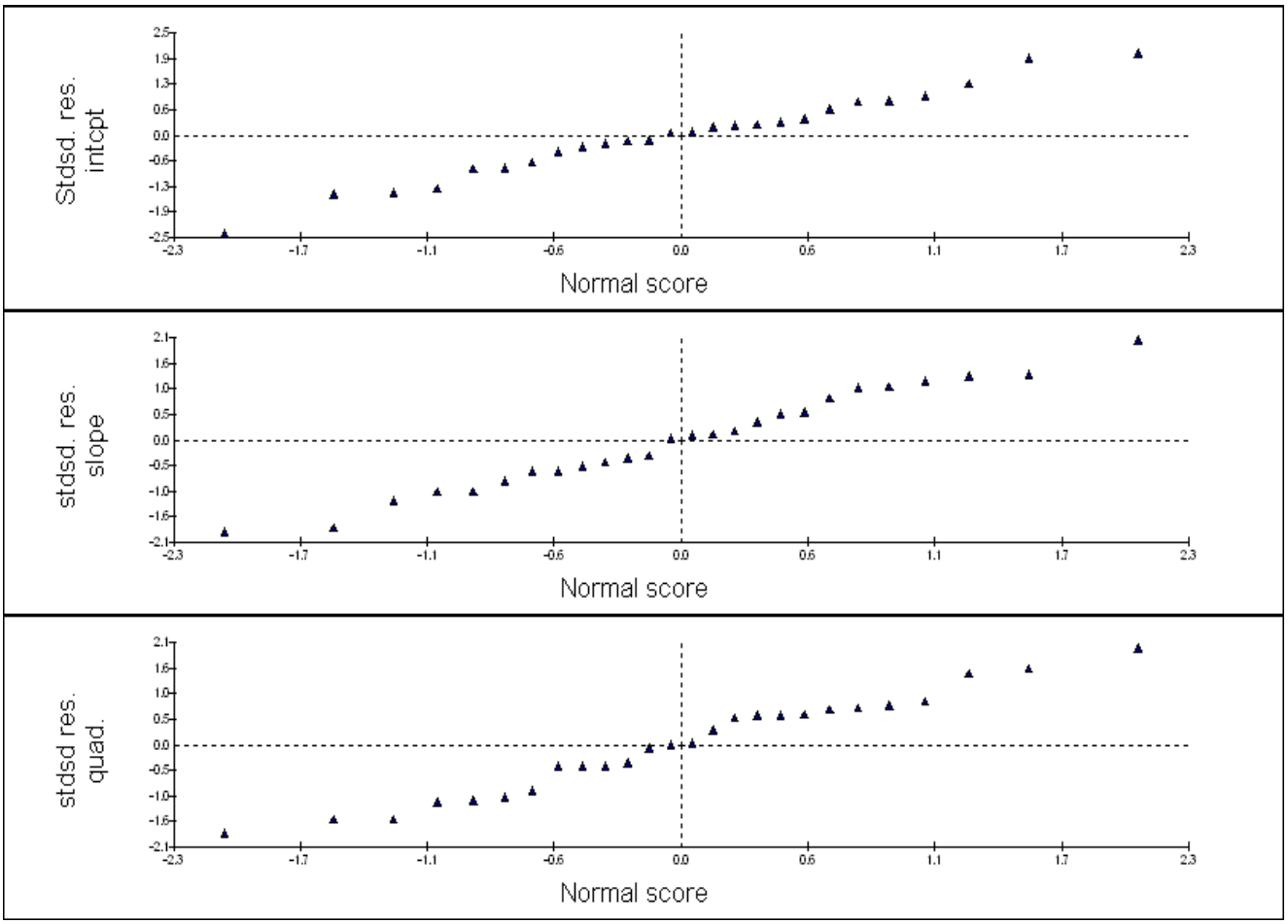


Figure 5. Standardised residuals at level 2 for model in column 3 in Table 7 with additional random effect for grafted quadratic.



7. Conclusions

In this tutorial we have described the fitting of models to repeated measures data using a multilevel model formulation that allows considerable flexibility and is easily generalisable. Our basic model can be extended to incorporate more complex data structures such as further levels of nesting, for example of individuals within schools or neighbourhoods or clinics, and to cross classifications where individuals are simultaneously classified, for example by where they live and where they are educated. We can also add further predictors such as an individuals' gender or ethnic background.

The use of regression splines further extends the usefulness of these models by allowing different degrees of polynomial to define the prediction curve over different time periods.

We have provided some historical background. Prior to the mid 1980s almost all the published analyses used multivariate approaches and these can still sometimes be found in contemporary literature, but are effectively best viewed as special cases of the more general approach using multilevel models. Thus, in the special case where there is a fixed number of discrete occasions the 2-level model becomes equivalent to the multivariate model. It is, however, more general and more flexible, especially since it easily allows further levels of nesting structure. With the ready availability of multilevel software, it can be recommended as the approach of choice for most purposes.

References

Bock RD. (1989) *Measurement of human variation: a two stage model. Multilevel analysis of educational data.* R. D. Bock. New York, Academic Press.

Goldstein H, Healy M JR and Rasbash J. (1994) Multilevel time series models with applications to repeated measures data: *Statistics in Medicine* 13, 1643-1655

Goldstein H. (1979) *The design and analysis of longitudinal studies*. London, Academic Press.

Goldstein H. (2003) *Multilevel statistical models*. London, Arnold.

Green PJ, Silverman BW. (1993) *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall/Crc monographs on statistics & applied probability)

Grizzle JC and Allen DM. (1969) An analysis of growth and dose response curves. *Biometrics* 25: 357-61.

Muthen B. (1997) Latent variable modelling of longitudinal and multilevel data. *Sociological methodology* 27: 453-480.

Pan H and Goldstein H. (1998) Multi-level repeated measures growth modelling using extended spline functions. *Statistics in Medicine* Vol. 17, 2755-2770

Rao CR. (1965) The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika,* 52, 447-458.

Tanner J. (1979) *A history of the study of human growth*. Cambridge University Press, Cambridge.

Wang Y. (1998) Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society*, B 60: 159-174.