

## THEORY AND METHODS

# Tutorial in Biostatistics: Evaluating the impact of 'critical periods' in longitudinal studies of growth using piecewise mixed effects models

Elena N Naumova,<sup>a</sup> Aviva Must<sup>a</sup> and Nan M Laird<sup>b</sup>

Recent developments in modern multivariate methods provide applied researchers with the means to address many important research questions that arise in studies with repeated measures data collected on individuals over time. One such area of applied research is focused on studying change associated with some event or critical period in human development.

This tutorial deals with the use of the general linear mixed model for regression analysis of correlated data with a two-piece linear function of time corresponding to the pre- and post-event trends. The model assumes a continuous outcome is linearly related to a set of explanatory variables, but allows for the trend after the event to be different from the trend before it. This task can be accomplished using a piecewise linear random effects model for longitudinal data where the response depends upon time of the event.

A detailed example that examines the influence of menarche on changes in body fat accretion will be presented using data from a prospective study of 162 girls measured annually from approximately age 10 until 4 years post menarche.

**Keywords** Growth, fat accretion, menarche, obesity, critical periods, longitudinal data, piecewise linear model, random effects model, mixed effects model

**Accepted** 12 March 2001

Evaluation of the impact of 'critical' or high-risk periods in longitudinal studies of growth may provide clues to the aetiology of life-long obesity or the optimal timing of preventive or therapeutic intervention. If the critical period has a lasting effect on an ongoing temporal process, we must be able to observe the changes in this temporal process with repeated measurements over pre- and post-critical periods. The ability to model the temporal pattern before and after the critical period provides the information needed to assess whether the critical period has had an enduring effect on the process under observation. The data that derive from such a study are inevitably and unavoidably complex because the timing of critical periods will vary from subject to subject and may not correspond to an occasion of measurement. In addition, subjects may exhibit considerable variability in time trends. Furthermore, the assumptions of independence required for traditional statistical methods

are not met in these studies due to the nature of the process itself and to the limitations of any observational study design.

Observational studies to assess the impact of critical periods share a number of challenging features. The critical or high-risk period is often a fairly complex biological process and it is sometimes difficult to define the beginning and end of the critical period. Pregnancy exemplifies a well-defined critical period: it starts with fertilization and ends with delivery. In this case, evaluation of the impact of pregnancy on fat accretion, for example, would be straightforward. We would compare fat accretion before and after pregnancy. When the critical period is less well defined and consists of a series of complex processes, one is forced to choose an event within the critical period, which will serve as a proxy. The event proxy allows the time period to be separated into pre- and post- event periods. The process of sexual maturation with menarche as the event is illustrative of a less well-defined critical period. Here, menarche is of interest to obesity researchers as a proxy for the pubertal period, proposed as a critical period in the development of obesity.

A second challenge arises because the critical period, and consequently the period around the event, is often characterized by increased variability. Increased variability arises for two major

<sup>a</sup> Tufts University School of Medicine, Department of Family Medicine and Community Health, Boston, MA, USA.

<sup>b</sup> Harvard School of Public Health, Department of Biostatistics.

Correspondence: Elena N. Naumova, Department of Family Medicine and Community Health, Tufts University School of Medicine, 136 Harrison Avenue, Boston MA 02111, USA. E-mail: elena.naumova@tufts.edu

reasons: first, the critical period, and the event *per se* can be thought of as a transient period of destabilization with uncertainty as to the timing of the event; second, since one does not know *a priori* when the event will occur, measurements at this time are not guaranteed. Menarche, for example, is characterized by changes in the levels of many hormones, some of which affect growth. Ideally, we would like a measurement at the time of the event, or, to increase the frequency of measurements just prior to the event. This, of course, requires one to know when the event will occur—the very information one lacks! Variability should be considered in assessing the impact of the event on the ongoing process.

A third challenge is the unbalanced design associated with the conduct of observational studies. In classically designed experiments, such as clinical trials or laboratory experiments, data can be collected completely and at standard intervals. Under these circumstances of balanced and complete data, covariates typically vary either within subjects (subject-specific level) or between subjects (population level), but not both. In observational studies, this is virtually never the case—unbalanced designs and/or missing data are the norm; it is rarely possible to design the study so that each subject is measured at the same set of times and measured at every single time. This unavoidable feature of observational studies should be incorporated into the analysis.

From a statistical perspective, to evaluate the impact of critical periods in longitudinal studies using a traditional analysis, one could use ordinary least squares (OLS) regression. One would need to do two regressions, one regression for the pre-event period and another for the post-event period, for each subject separately. Then, taking the difference of slopes for each subject, one could calculate the average difference in the slope pairs, its standard error, and perform a *t*-test to test whether this value is different from zero. The advantage of this approach is that it is simple to do and easy to explain. However, this approach does not take full advantage of the richness of the data, including the correlation structure. Also, the traditional analysis does not characterize individual variation relative to the population mean. In addition, this approach requires one to discard data on subjects with an insufficient number of measurements either before or after the event. This approach will also weight all remaining subjects equally even though their estimates are based on variable numbers of measurements. Finally, fitting two regressions separately will not insure that the lines will meet at the time of event, which is required by the growth process. It is not biologically possible to have an substantial instantaneous shift in body fat.

These limitations are elegantly addressed by modern applied statistical techniques, such as general linear mixed models.<sup>1</sup> The mixed model intrinsically consists of two components: subject-specific and population, while accounting for the correlation at each level. The model allows one to characterize individual variation relative to the population mean. The estimates that derive from a mixed model are weighted based on the actual data. Missing data can be handled readily by these techniques. Finally, the mixed model allows one to simultaneously fit the two slopes, so that intersection of the lines at the time of the event is insured.

Given the clear advantages for the applied researcher, why have these methods not been utilized more widely? There are likely several reasons for this. One reason may be that unless compelled to do otherwise, researchers will continue to use

familiar methods. The typical researcher may be intimidated by complex formulations and mathematical notation (e.g. Greek letters, double subscripts, matrix notation, multiple brackets). In addition, the complete absence of graphical displays, which can make the step-by-step process far easier to understand, are not often included. Papers describing these methods typically appear in statistical journals and, although they include clear examples and often do explain the model in terms of the variables under study, the amount of mathematical terminology the reader must wade through before they get to interpretation may be overwhelming. These descriptions meet the needs of the statistical reader, but may be difficult for the applied researcher.<sup>2</sup>

Another issue is terminology: the literature on repeated measurement analysis describes the same models and methods using different terminology, and the lack of consistency is a source of confusion. For example, random effects, mixed effects, two-stage and hierarchical models are variously used to describe the same model. The term hierarchical is especially common in the educational literature, where one may have repeat measures on children in several classrooms in several schools, etc. In this paper, we use the term random effects to describe models where all the parameters in a single individual's regression model are random. We use the term mixed effects to describe models where the parameters can be both fixed and random. Incorporating a fixed effect in the model allows one to include covariates that do not vary within a subject. As another example, consider the interchangeable usage of the terms repeated measures and longitudinal data. Sometimes the term repeated measures is used narrowly to refer to designs where each subject is measured under different conditions, and then they are appropriately analysed with a classical repeated measures analysis. In other instances, the term is used more broadly to include longitudinal studies, where the repeated measures are taken on the same subject over time. To our minds, longitudinal data represent a special case of repeated measures, where the condition is the time at which the measurement was obtained.

Addressing this need is particularly urgent. With the number of large longitudinal studies currently underway, these data will become more common, and they deserve to be analysed properly. The software needed for analysis of longitudinal data obtained from complex study designs is now part of standard statistical packages.<sup>3</sup> User-friendly graphic interfaces in the commercially available software allow one to easily visualize each step of data analysis, validation and presentation of results. Together these developments suggest the opportunity to reach a broader audience.

Our goal is to illustrate and present the use of general linear mixed model in a manner that is accessible to applied researchers; we use prospectively collected data to assess the influence of menarche on changes in body fat accretion as an example. We take the reader through the *process* of data preparation, model formulation, fitting and validation. In this process, we will incorporate traditional and familiar statistical methods, as needed. To visualize the hypothesis under study, actual data are used, and, for practically every step in the development of the mixed model, we incorporate a variety of graphical displays.

## Description of the Study

It is known that the peripubertal period is associated with increases in body fatness.<sup>4</sup> This increase starts just before or

around menarche and is presumed to level off by approximately 4 years after menarche. However, this phenomenon has not been systematically studied in population-based samples. The current epidemic of obesity in the US and elsewhere has stimulated interest in the aetiology, with the possible role of critical periods in the development of obesity as one important focus of efforts.<sup>5,6</sup>

The data presented here were derived from a prospective study of the development of obesity in a cohort of 162 girls. Initiated in 1990, this ongoing NIH-funded study examines the effects of energy expenditure and other metabolic and behavioural factors on subsequent changes in body fat as adolescence progresses, until 4 years after menarche. At study entry, all girls were non-obese, as determined by a triceps skinfold thickness less than the 85th percentile, and pre-menarcheal. All girls were followed according to a schedule of annual measurements. They reported their date of menarche and were then observed until their exit day, which was scheduled on the fourth anniversary of their menarche.

The outcome variable of interest in the study, body fatness, is based on bioelectric impedance analysis. Per cent body fat (%body fat) was obtained from three basic measurements: weight (Wt [kg]), height (Ht [cm]) and bioelectric impedance resistance (R). Per cent body fat was then calculated using the Kushner equation:<sup>7</sup> % BF =  $(1 - TBW/0.73 \text{ Wt}) \times 100\%$ , where total body water  $TBW = (0.7 \text{ Ht}^2/R) - 0.32$ . For each subject, date of birth, date of study entry, dates of each follow-up measurement, date of menarche, and date of exit were recorded. Then, for baseline and subsequent measurements, the age and times relative to menarche and corresponding age at the date of each measurement were calculated in years, to three significant digits after the decimal point.

Although the endpoint of the study is 4 years after the first menstrual period (menarche), the study is ongoing. As of June 1999, based on protocol completion, of the 162 girls enrolled in the study and followed through menarche, 94 girls have exited (they are the completers), 54 girls remain in the study, and 14 are lost-to-follow-up after menarche. On average, the time period

between two consecutive measurements was  $1.11 \pm 0.3$  years. There are 1049 individual %body fat measurements, with an average of  $6.5 \pm 1.41$  measurements per subject:  $3.06 \pm 1.36$  for the pre-menarcheal period (497 measurements) and  $3.45 \pm 1.22$  for the post-menarcheal period (552 measurements). In our data, the number of measurements per subject pre- and post-menarche are almost equal; Table 1 provides the overall frequency of measurements.

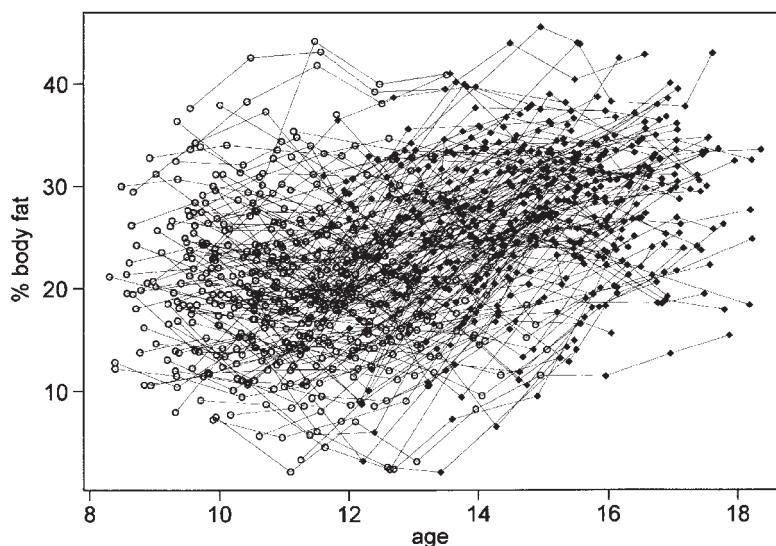
## Exploratory Analysis

We conducted an exploratory analysis to check for the presence of unusual features in the data, to visualize the hypothesis under study, and to check the validity of our initial assumptions for modelling. Per cent body fat is shown as a set of individual growth curves. Figure 1 shows the entire set of data, in which each data point is shown as an empty circle for pre-event and a solid diamond for post-event measurements. This plot reveals the variability in subjects' age at entry, age at exit and age at the menarcheal event, but includes too much data to reveal the underlying patterns.

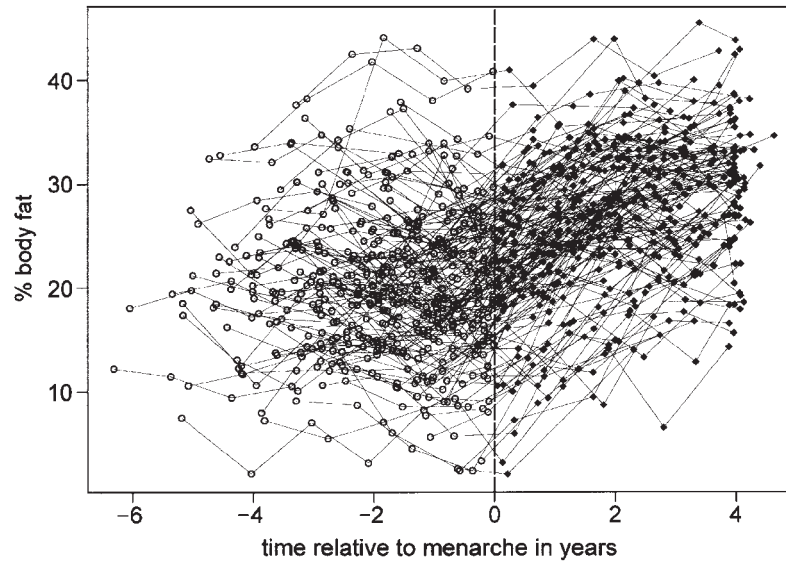
Rescaling the horizontal axis of this graph more clearly shows the pattern of growth. Figure 2 represents the same set of data,

**Table 1** Overall frequency of measurements

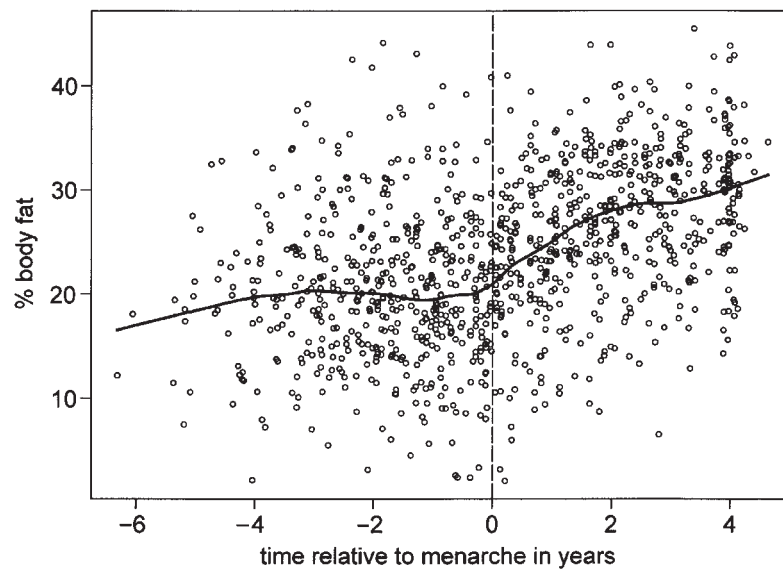
No. of measurements per subject	No. of subjects (%)	Total no. of measurements (%)
3	7 (4)	21 (2)
4	3 (2)	12 (1)
5	29 (18)	145 (14)
6	39 (24)	234 (22)
7	45 (28)	315 (30)
8	30 (19)	240 (23)
9	8 (5)	72 (7)
10	1 (1)	10 (1)
<b>Total</b>	162	1049



**Figure 1** The entire set of growth curves, where each data point is shown as an empty circle for pre-menarcheal and as a solid diamond for post-menarcheal measurement



**Figure 2** The entire set of growth curves plotted against the time relative to menarche. Each data point is shown as an empty circle for pre-menarcheal and as a solid diamond for post-menarcheal measurement



**Figure 3** Per cent body fat measurements with LOWESS curve non-parametrically smoothed with a moving window chosen to cover 20% of data points (overlay)

but for a time scale relative to the individual age at menarche. In this graph we can see more clearly the separate patterns before and after the menarcheal event. Also, we can see that the range of measurement times in the pre-menarcheal period is greater than that in the post-menarcheal range, which was limited to 4 years by the study design.

To reveal the general pattern in growth curves with respect to the time of event, we smooth the outcome variable using non-parametric robust local smoothing procedures.<sup>8</sup> Although this technique ignores the correlation structure in repeated measurements, it allows for an informal check for non-linearity

in the growth pattern. A scatterplot of %body fat and the results of smoothing are shown in Figure 3. One can see that the smoothed curve remains flat during the time period prior to menarche, and then starts to rise sharply near the time of the event, remains steep for about 2 years after menarche and then becomes shallower.

As the part of the preliminary analysis, we evaluated the descriptive statistics for the cross-sectional examination of %body fat in pre-menarcheal and post-menarcheal periods. We also calculated the summary statistics for differences in subject's post- and pre-averages (Table 2). Checking for departures from

**Table 2** Descriptive statistics for cross-sectional examination of per cent body fat

Time period	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Var.	SD
All period	2.112	18.24	23.59	23.61	29.56	45.49	62.95	7.93
Pre-menarcheal	2.428	15.41	20.35	20.34	24.26	40.60	48.20	6.94
Post-menarcheal	8.043	22.76	26.38	26.35	31.46	41.63	39.60	6.29
Post-pre difference	-12.29	3.536	6.607	6.132	9.249	17.51	18.48	4.30

normality in subject's post- and pre-averages using standard tests, we did not observe any obvious departures. However a slight right skew in pre-event and some left skew in post-event measurements was observed.

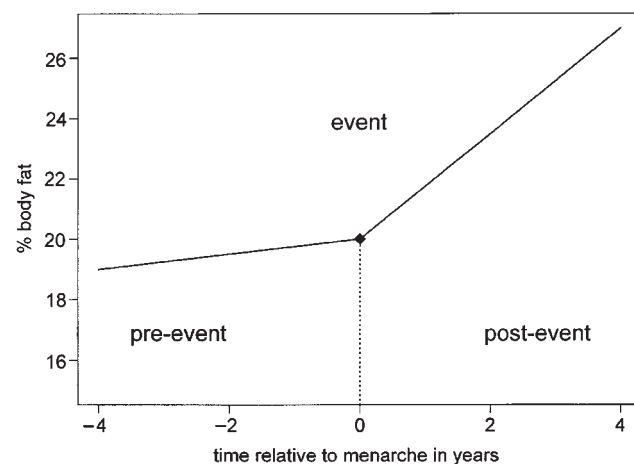
## Research Hypothesis

We hypothesize that %body fat accretion increases linearly with age, but with different slopes before and after menarche. We assume that each subject has a two-piece linear spline growth curve with a knot at the time of menarche. Graphical representation of the research hypothesis and the modelling approach are shown schematically in Figure 4. To make the graph realistic and to reflect the range of time of measurements relative to the age at menarche, we used results from our exploratory analysis on sample means of age at baseline (10 years), of age on exit day (16 years), average age at menarche (12.8 years), and average of %body fat taken at the baseline and exit.

Our goal is to define a model that allows us to directly quantify the scientific question, i.e. to evaluate the two slopes and their difference. In addition, the model should readily allow testing of secondary hypotheses, and to permit model validation. Thirdly, the model should be easily fit using standard statistical software packages.

## Piecewise Least Squares Analysis

For the piecewise least squares analysis, we fit two straight lines to the data for each girl, connected at the time of menarche. We regress %body fat as a function of time relative to menarche, obtaining one intercept and two slopes: one for before and one



**Figure 4** Graphical representation of the research hypothesis and modelling approach: the piecewise model assigns a pre-event slope  $\beta_1$  and a post-event slope  $\beta_2$  separated by the menarcheal event

for after the event. Then, we take the difference in slopes for each girl. Next, we compute the mean intercept, mean slopes, the mean slope differences, and their standard errors. Finally, we perform a paired *t*-test in the usual way. Note that for this piecewise least squares regression, to estimate the three regression parameters, we need at least three data points (with at least one data point for both pre- and post- menarcheal time period), and known age at menarche. The number of girls with sufficient data to calculate the intercept and two slopes is 153.

The mean and standard deviation estimates are as follows: for pre-menarcheal slope:  $0.81 \pm 4.46$ ; for post-menarcheal slope:  $2.44 \pm 1.93$ ; and for post- and pre- difference:  $1.62 \pm 5.17$ . The correlation between slopes is weak and inverse ( $r = -0.18$ ,  $P = 0.027$ ). Since we are regressing %body fat as a function of time relative to menarche, the mean and standard error of the intercept provides the estimate of %body fat at menarche:  $21.41 \pm 7.12\%$ . T-tests demonstrate that the post-, but not the pre-menarcheal slope differs significantly from zero, as does the difference in the two slopes.

Although this piecewise regression is simple to compute and interpret, there are a few disadvantages in this procedure: first, girls with fewer measurements are weighted equally to girls with more measurements in our estimation of population mean, and second, we must disregard data on subjects with insufficient number of data points and therefore reduce statistical power.

## Random Effects Analysis

Here we provide a description of a piecewise approach for modelling growth using a random effects analysis. This analysis assumes that a continuous outcome variable is linearly related to a set of explanatory variables, allows one to fit pre- and post-slopes simultaneously using all data points, and also can account for dependence between observations collected over time.

We assume that our measurements depend upon the menarcheal event such that each subject has a baseline level, a pre-event slope and a post-event slope. Let  $d_i$  be age at menarche for *i*-subject and  $a_{ij}$  be age at *j*-measurement for *i*-subject, then  $t_{ij} = a_{ij} - d_i$  is time of *j*-measurement for *i*-subject before or after menarche. Let  $\delta_{ij}$  be an indicator:  $\delta_{ij} = 1$  for the time period before menarche,  $t_{ij} < 0$ , and  $\delta_{ij} = 0$  for the time period after menarche  $t_{ij} \geq 0$ . Now, we specify a model at the subject specific level for the *n* subjects  $i = 1, \dots, n$ , each of whom has  $m_i$  observations,  $Y_{ij}$ ,  $j = 1, \dots, m_i$  so that for the time period before menarche  $Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$  ( $\delta_{ij} = 1$ ), for the time period after menarche  $Y_{ij} = \beta_{0i} + \beta_{2i}t_{ij} + e_{ij}$  ( $\delta_{ij} = 0$ ), and the combined model is:

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij}\delta_{ij} + \beta_{2i}t_{ij}(1 - \delta_{ij}) + e_{ij}, \quad (1)$$

where  $(\beta_{0i}, \beta_{1i}, \beta_{2i})$  are the intercept and two slopes for the *i*<sup>th</sup> subject,  $t_{ij}$  is time of measurement relative to menarche, and

$e_{ij}$  is the measurement error at the  $j^{\text{th}}$  occasion for the  $i^{\text{th}}$  subject. These errors,  $e_{ij}$ , are typically taken to be independently and identically normally distributed with a mean  $E(e_{ij}) = 0$  and variance,  $\text{var}(e_{ij}) = \sigma^2$ . When the underlying biological process for each subject is adequately represented by the random effects model, the correlation among the  $e_{ij}$  should be negligible (meaning that there is no association between the residuals within a single subject after subtracting out individual growth curves), and the  $e_{ij}$  can be thought of as measurement or sampling error, and  $\sigma^2$  as the within-subject variance.

At the population level, the model specifies the population mean and variance:

$$E \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \text{var} \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{bmatrix} \sigma_{00}^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_{11}^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$$

These matrices have a straightforward interpretation:  $(\mu_0, \sigma_{00})$ ,  $(\mu_1, \sigma_{11})$ , and  $(\mu_2, \sigma_{22})$  are respectively the mean and standard deviation: for ‘true’ %body fat at menarche; of the pre-event slopes; and of the post-event slopes respectively. Similarly,  $\sigma_{12}$ ,  $\sigma_{01}$  and  $\sigma_{02}$  are: the covariance of the pre-event and post-event slopes; the covariance of ‘true’ %body fat at menarche and the pre-event slopes; and the covariance of ‘true’ %body fat at menarche and the post-event slopes. Because the actual %body fat at menarche is not observed and not directly estimable, we use the term ‘true’ %body fat at menarche to emphasize that this  $\beta_{0i}$  is a parameter in the piecewise linear model.

If we assume that  $(e_{ij}, e_{ik}; j \neq k)$  are independent (residuals are measurement errors), then the overall variances and correlations depend upon the actual times of the observations, and upon population variances and covariances. Note that at the subject-specific level  $\sigma^2$  equals  $\text{var}(e_{ij})$  and represents the measurement error. Therefore, for the time period before menarche, variances and covariances have the following form:  $\text{var}(Y_{ij}) = \sigma^2 + \sigma_{00}^2 + \sigma_{11}^2(t_{ij})^2 + 2t_{ij}\sigma_{01}$  and  $\text{cov}(Y_{i1}, Y_{i2}) = \sigma_{00}^2 + \sigma_{11}^2(t_{i1}t_{i2}) + \sigma_{01}(t_{i1} + t_{i2})$ . Similarly for the time period after menarche:  $\text{var}(Y_{ij}) = \sigma^2 + \sigma_{00}^2 + \sigma_{22}^2(t_{ij})^2 + 2t_{ij}\sigma_{02}$  and  $\text{cov}(Y_{ik+1}, Y_{ik+2}) = \sigma_{00}^2 + \sigma_{22}^2(t_{ik+1}t_{ik+2}) + \sigma_{02}(t_{ik+1} + t_{ik+2})$ .

The random effects model provides users with the means to illustrate and make estimates for individual curves. The population estimates:  $\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2$ , and individual estimates:  $\hat{\beta}_{0i}, \hat{\beta}_{1i}, \hat{\beta}_{2i}$  can be used to create individual predicted trajectories, or growth curves. This feature of the model can be used, in theory, to assess how an individual girl’s fat accretion differs or mirrors that of the population generally. How well this will work in practice depends on the number of individual measurements upon which a girl’s curve is based. This has implications for practical study design issues, where with fixed resources, one often has to balance the number of subjects against the number of measurements per subject.

The model parameters can be estimated using maximum likelihood (ML) or its variant, restricted maximum likelihood (REML), assuming that the observations are normally distributed. The ML estimates of the mean are in essence weighted least squares, with weights calculated from the estimated variances. The REML is usually preferred for estimation of the variance-covariance components because it adjusts for the degrees-of-freedom lost in estimating the means. Parameter estimates, and their standard

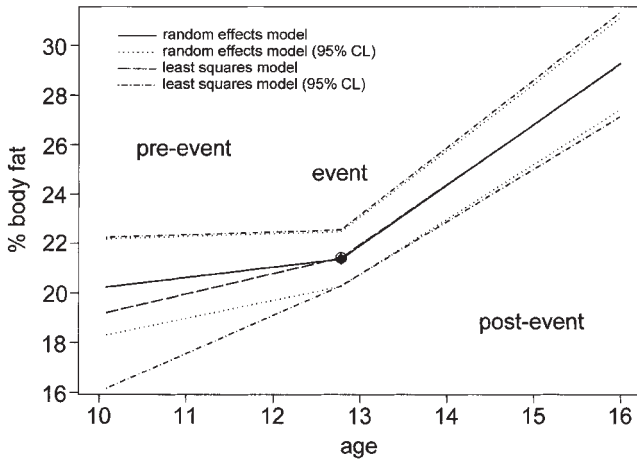
errors, can be obtained using a variety of software packages, including SAS (Proc MIXED), S-Plus (lme) and BMDP-5V.

We now apply the piecewise random effects model to estimate the pre- and post-menarcheal slopes and to predict body fatness. The results of this model consist of two main parts: a set of individual intercepts and two slopes (random effects), and estimates of their means (fixed effects) and standard deviations. Also, the correlations between the slopes and intercepts provide additional useful information. Even though the results for the individual curves form the basis for the fixed effects estimation, for pedagogical reasons, our description of results will begin with the fixed effects, since these results address our primary research question.

Our goal is to assess the influence of the menarcheal event on the rate of fat accretion. The fixed effects results indicate that the mean and standard deviation of the pre-menarcheal slope, the post-menarcheal slope, and %body fat at menarche are:  $0.42 \pm 1.28 (\hat{\mu}_1, \hat{\sigma}_{11})$ ;  $2.46 \pm 0.94 (\hat{\mu}_2, \hat{\sigma}_{22})$ ; and  $21.36 \pm 6.78 (\mu_0, \sigma_{00})$ , respectively. As a next step, we use the results of the mixed effects model to test our hypotheses. First, we formally test the hypothesis: Is %body fat increasing before and after menarche? The estimate for the population mean pre-menarcheal slope ( $\hat{\mu}_1 \pm \text{SE}$ ) is  $0.42 \pm 0.16$ , with  $Z = 2.66$ , which is statistically significant (using the Z-test). The slope is quite shallow (the estimated rate of fat accretion is 0.42%), but different than zero. The high variability suggests that some girls are gaining fatness and others are losing fatness over the pre-menarcheal period. After menarche, the estimate for the population mean post-menarcheal slope ( $\hat{\mu}_2 \pm \text{SE}$ ) is  $2.46 \pm 0.12$ , and  $Z = 20.67$ , which translates to %body fat of 21.4% with a 95% CI ranging from 8.08% to 34.64%. Our focus now shifts to the second hypotheses regarding the difference in the pre- and post- menarcheal slopes:  $\hat{\mu}_2 - \hat{\mu}_1$ . That difference is  $2.05 \pm 0.22$ , with Z for the hypothesis that slopes are equal of 9.32. Fat accretion is more uniform after menarche, and the rate of fat accretion is almost six times greater than before menarche. So for each year of age after menarche %body fat increases an average 2.46%, so that by 4 years post-menarche, body fatness has increased by approximately 10% (absolute %body fat). If we assume that the individual slopes are approximately normally distributed in the population, then, based on the results of the random effects model, about 37% of girls have pre-menarcheal slopes  $< 0$ , and virtually none have post-menarcheal slopes  $< 0$ . We conclude that the post-menarcheal slope is significantly greater than the pre-menarcheal slope.

The random effects model also provides very useful additional information, such as an estimate of the measurement error ( $\hat{\sigma} = 3.1$ ) and covariance estimates and correlations: the correlation of %body fat at menarche and the pre-event slope ( $\hat{\rho}_{01} = 0.29$ ), the correlation of %body fat at menarche and the post-event slope ( $\hat{\rho}_{02} = -0.56$ ), and the correlation of the pre- and post-event slope ( $\hat{\rho}_{12} = -0.10$ ).

A graphical display of the resulting curve obtained from the random effects model and least squares model along with their confidence intervals are shown in Figure 5. The graph is rescaled to reflect the range of predicted %body fat at the expected average age at menarche (12.8 years), and for the time period of 4 years before and after menarche. In both models, the predicted fat accretion before menarche is unimpressive, but after menarche %body fat increases markedly. Although the predicted



**Figure 5** Graphical representation of the modelling results: predicted %body fat along with the 95% upper and lower confidence limits, based on the piecewise random effects model (the solid line—for fixed effects, the dotted lines—for the confidence interval, and the solid diamond—for predicted %body fat at menarche) and least squares model (the dashed line—for fixed effects, the dotted-dashed lines—for the confidence interval, and the open circle—for predicted %body fat at menarche)

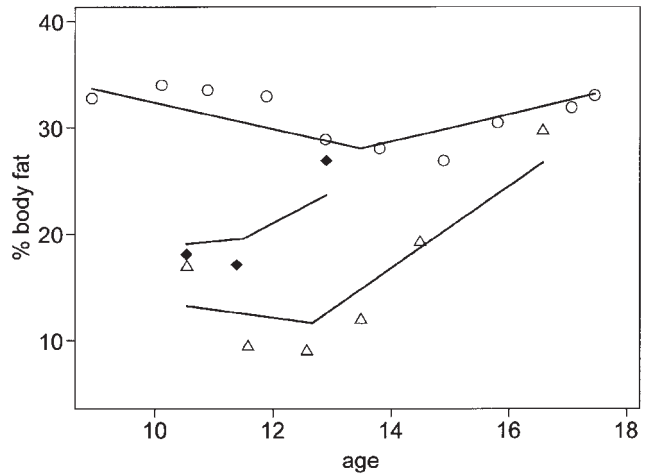
%body fat at menarche and post-menarcheal slope are very close in both models, the confidence intervals for the pre- and post-event slopes are narrower for random effects model estimates than for the least squares analysis which results from the weighting of the random effects estimates. Thus, the random effects model reduces variability and thereby improves the efficiency of estimation.

Next, we turn our attention to the individual growth curves, where there are two aspects of interest: the individual curves themselves and their correlation structure. Figure 6 provides an example of ‘reconstructed’ individual curves for three girls (with 10, 6 and 3 measurements) using the random effects estimates. Again, weighting of random effects estimates ensures that subjects with more data points have curves which are closer to their data, while subjects with fewer data points have curves closer to the population mean.

### General linear mixed model: adding model complexity

The random effects model represents an early step in data exploration. However, the researcher is often interested in what variables influence changes over time. For the present example, there are several questions that are of potential interest. If adolescence does in fact represent a critical period, we are interested in variables that might influence %body fat at menarche and the post-event (post-menarcheal) slope. Among the many questions one might pose are: Is there non-linearity in the post-menarcheal time period? Are there any variables measured at baseline that influence %body fat at menarche and/or the magnitude of the slopes? Examples of potential variables include ethnicity, socio-economic status, or parental obesity.

To answer these questions, we extend our random effects model to a mixed effects model to make it more flexible. In the linear mixed effects model, fixed effects and random effects



**Figure 6** Individual curves for three girls

are connected to each other, so that in a single model for  $Y_{ij}$  any observable effect is a combination of the two. To see how mixed effects are formulated, let  $b_{0i} = \beta_{0i} - \mu_0$ ;  $b_{1i} = \beta_{1i} - \mu_1$ ;  $b_{2i} = \beta_{2i} - \mu_2$ , so  $E(\mathbf{b}_i) = 0$  and  $var(\mathbf{b}_i) = var(\beta_i)$ . Then, the piecewise random effects model can be rewritten as follows:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} \delta_{ij} + \beta_2 t_{ij} (1 - \delta_{ij}) + b_{0i} + b_{1i} t_{ij} \delta_{ij} + b_{2i} t_{ij} (1 - \delta_{ij}) + e_{ij}, \quad (2)$$

where  $t_{ij} \delta_{ij}$  and  $t_{ij} (1 - \delta_{ij})$  are the pre- and post- event times, and the population mean effects are now denoted  $\beta_0, \beta_1, \beta_2$ . This reformulation provides a high level of flexibility. It allows us to model the systematic variation in the data which can be linked to explanatory variables that differ among the subjects (in fixed effects:  $\beta_0, \beta_1, \beta_2$ ) and that vary within the subject (random effects:  $b_{0i}, b_{1i}, b_{2i}$ ).

Thus, to answer the questions posited above we will extend the model by adding parameters as fixed effects and/or random effects. To answer the first question regarding non-linearity, we can add a quadratic term to the model, such as  $t_{ij}^2 (1 - \delta_{ij})$ . To answer the second question we can add baseline variables, which influence the %body fat at menarche and possibly the slopes as well. To evaluate an effect of a baseline variable on an intercept, the baseline variable should be added as a main term. To evaluate an effect of a baseline variable on a slope we should add an interaction term for this variable and time since menarche. If the baseline variable does not change over time, we are adding it to the model by extending the fixed effects terms.

Here we provide the results of the modelling from three mixed effects models. Model A includes the quadratic term for the post-menarcheal slope as fixed and random effects and is formulated as follows:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} \delta_{ij} + \beta_2 t_{ij} (1 - \delta_{ij}) + \beta_3 t_{ij}^2 (1 - \delta_{ij}) + b_{0i} + b_{1i} t_{ij} \delta_{ij} + b_{2i} t_{ij} (1 - \delta_{ij}) + b_{3i} t_{ij}^2 (1 - \delta_{ij}) + e_{ij}. \quad (3)$$

We expect this model to improve the fit of the post-menarcheal data due to potential for non-linearity.

Model B includes stratification by parental obesity status as a fixed effect (where  $v = \{0, 1 \text{ or } 2\}$ , if none, one, or both parents were obese) and is formulated as follows:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} \delta_{ij} + \beta_2 t_{ij} (1 - \delta_{ij}) + \beta_3 v_i + b_{0i} + b_{1i} t_{ij} \delta_{ij} + b_{2i} t_{ij} (1 - \delta_{ij}) + e_{ij}. \tag{4}$$

This model includes a variable collected at baseline reflecting parental obesity, which is posited to influence model parameters. In order to assess the influence of parental obesity on %body fat at menarche,  $\beta_3 v_i$  is added as a fixed effect. Model B only tells us how the number of obese parents affects the %body fat at menarche.

Model C examines the influence of the parental obesity status on the post-menarcheal slope by the addition of an interaction term:  $\beta_4 t_{ij} (1 - \delta_{ij}) v_i$ . The model is formulated as follows:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} \delta_{ij} + \beta_2 t_{ij} (1 - \delta_{ij}) + \beta_3 v_i + \beta_4 t_{ij} (1 - \delta_{ij}) v_i + b_{0i} + b_{1i} t_{ij} \delta_{ij} + b_{2i} t_{ij} (1 - \delta_{ij}) + e_{ij}. \tag{5}$$

The estimates of the fixed effects from the basic random effects model and Models A, B and C are shown in Table 3. All models produce very similar estimates for the intercept, %body fat at menarche. The results of Model A suggest that there is a significant non-linearity in the post-event slope, which tends to level off by 4 years post-menarche. Model A also produces the smallest estimate for the pre-event slope. Model B suggests only a weak positive effect of parental obesity on body fat at menarche. The results of Model C suggest that parental obesity has no appreciable effect on the post-menarcheal slope. The piecewise nature of these models results in a strong interdependence among the parameters, so that a change in the intercept, for example, is reflected in changes in the two slope estimates.

### Summary

This paper serves a dual purpose: to explain the advantages of using mixed effects models for longitudinal growth data and to demonstrate the model's utility in answering a question of epidemiological interest. The current thinking regarding the natural history of obesity is characterized by the invocation of critical periods. Because these mixed effects models can accommodate a change in slope, they are particularly well-suited for modelling growth data wherein such events occur. In regards to critical periods, the results of the models presented are quite clear and confirm what has been observed impressionistically. Clinical

observation suggests that for many girls the first few years post-menarche are associated with appreciable increases in fatness.<sup>4</sup> Our findings indicate that adolescence is a time of potential increased obesity incidence and that menarche does, in fact, emerge as the breakpoint in the fat accretion curve. In our analyses the inflection point is observed to occur (Figure 3), and the timing of that point of inflection turns out to be around menarche. Due to the absence of measurement at the time of menarche, the interval around this event represents a time of uncertainty. Using the approach of regressing %body fat against time relative to menarche we have acquired from this model the information, which is practically unobservable in a field study, the %body fat at menarche.

Use of random and mixed effects models substantially reduces parameter variability for all three estimates and thereby improves the efficiency of estimation. For pre- and post-menarcheal slopes, the standard deviations are 3.5 and 2 times smaller for random effects model estimates than for the least squares analysis. Although the large variation in estimated slopes might be explained by inherent biological variability during the critical period of sexual maturation, the large variation in estimated slopes obtained from OLS can also be exaggerated for at least two reasons: first, girls with fewer measurements are weighted equally to girls with more measurements in the estimation of population mean slope; and secondly, slope estimates depend on the number of measurements available. The requirement to have three data points and age at menarche to estimate both slopes and intercept in OLS would omit all subjects who were not yet menarcheal even if they have enough data points for pre-menarcheal time period. This clearly reduces the efficiency of OLS model for the interim and overall analysis.

Some practical considerations are worthy of mention. Because the mixed effects models are basically weighted least squares, with the weights estimated from the data, and because hypothesis testing relies on asymptotic normality of the estimated fixed effects, one needs a reasonable (>30) number of subjects to ensure that this property holds. Secondly, given that missing responses and attrition will occur in practice, the validity of the results depends upon the assumption that missingness is not related to unobserved measurements. The research question itself, the exploration of the pre- and post-menarcheal slope differences, emphasizes the importance of subject retention in a longitudinal study. Good retention ensures that the reasons for leaving a longitudinal study are likely to be largely independent of growth pattern, so that missingness assumptions are satisfied. Although subjects without a known time of menarche are

**Table 3** The estimates of the fixed effects from random effects model and Models A, B and C

Parameter	Random Effects		Model A		Model B		Model C	
	Estimate ± SE	z ratio	Estimate ± SE	z ratio	Estimate ± SE	z ratio	Estimate ± SE	z ratio
Intercept	21.36 ± 0.565	37.84	20.43 ± 0.581	35.14	20.34 ± 0.765	26.57	20.42 ± 0.812	25.13
Pre-event slope	0.42 ± 0.157	2.66	-0.01 ± 0.161	-0.07	0.39 ± 0.160	2.44	0.39 ± 0.160	2.44
Post-event slope	2.46 ± 0.119	20.67	4.83 ± 0.318	15.15	2.46 ± 0.123	19.92	2.43 ± 0.175	13.86
Quadratic term for post-event slope			-0.65 ± 0.075	-8.55				
Parental obesity					1.51 ± 0.677	2.23	1.40 ± 0.766	1.83
Interaction term for parental obesity and post-event slope							0.05 ± 0.16	0.29



excluded and this may bias the estimated pre-menarcheal slope, including these subjects may bias the estimates of change in slope. Our approach uses only within-subject data to estimate this change in slope.

Random effects and mixed effects models are very flexible and can accommodate a variety of study designs, data models and hypotheses. We have presented an example with a continuous Gaussian outcome variable; these models can readily be extendable to accommodate discrete outcomes as well. As described above, from a statistical perspective these approaches address many of the challenges present in longitudinal studies. The software needed to perform the analysis is readily available as part of standard statistical software programs. These programs have incorporated special procedures for handling missing data and have built-in

user-friendly graphical interfaces that permit the analyst to visualize the data at each step. The applied researcher would do well to learn how to use these methods today, if not sooner.

## Acknowledgements

We wish to acknowledge June Stevens, Ph.D., session organizer for our tutorial entitled, 'The Application of Longitudinal Analysis Methods to Repeated Measures of Body Weight and Fat Distribution' presented at the Annual Meeting of the North American Association for the Study of Obesity, Charleston, South Carolina, 14–18 November 1999. The data derive from a study supported by NIH grants DK-HD50537, DK46200 and M01-RR-00088.

### KEY MESSAGES

- Current growth research focuses on the evaluation of the impact of 'critical' or high-risk periods using longitudinal study designs.
- Piecewise regression can be used to model changes during critical periods.
- Regressing growth characteristics against time relative to the event provides estimation for parameters that are not observable in most human studies.
- Mixed effects models provide the ability to characterize individual variation relative to the population mean.
- The model suggests that fat accretion is different for pre- and post-menarche, with greater fat accretion following the critical event of menarche.

## References

- <sup>1</sup> Hogan JW, Laird NM. Intention-to-treat analyses for incomplete repeated measures data. *Biometrics* 1996;**52**:1002–17.
- <sup>2</sup> Cnaan A, Laird NM, Slasor P. Tutorial in biostatistics: using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Stat Med* 1997;**16**:2349–80.
- <sup>3</sup> Kreft IGG, deLeeuw J, van der Leeden R. Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *Am Statistician*. 1994;**48**:324–35.
- <sup>4</sup> Garn SM, LaVelle M. Two-decade follow-up of fatness in early childhood. *Am J Dis Child* 1985;**139**:181–85.
- <sup>5</sup> Allison D, Paultre F, Heymsfield SB, Pi-Sunyer FX. Is the intra-uterine period really a critical period for the development of adiposity? *Int J Obesity Relat Metab Disord* 1995;**19**:397–402.
- <sup>6</sup> Dietz WH. Critical periods in childhood for the development of obesity. *Am J Clin Nutr* 1994;**59**:955–59.
- <sup>7</sup> Kushner R, Schoeller D, Fjeld CR, Danford L. Is the impedance index (ht<sup>2</sup>/R) significant in predicting total body water? *Am J Clin Nutr* 1992;**56**:835–39.
- <sup>8</sup> Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Statist Assoc* 1979;**74**:829–36.
- <sup>9</sup> Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. New York: Oxford University Press, 1994.
- <sup>10</sup> Pinheiro JC, Bates DM. *Mixed-effects Models in S and S-PLUS*. New York: Springer-Verlag, 2000.

## Appendix

### Available software and recommendations for data preparation

To make this approach amenable to wide use, we developed a set of recommendations for step-by-step data preparation, exploratory analysis, model implementation, and visualization of results using standard statistical software. The interested reader is directed to previous reviews of available software, such as BMDP 5-V, SAS, S-Plus, HLM, ML3<sup>3,9</sup> and codes for implementing these procedures previously published in the statistical literature.<sup>2</sup> Pinheiro and Bates provide guidance for using S-Plus for mixed effects modelling with a strong emphasis on the use of graphical displays.<sup>10</sup>

In brief, to prepare the data set, one extracts data from baseline and annual follow-up files, checks data for completeness and develops decision rules for inclusion in the model. The data are then reformatted to fit the scheme for a covariate matrix, e.g. each row is a single record with identification number, outcome, time of measurement, and pre/post event index. The plots of the raw data for the complete and selected data set can be very useful in making assumptions for modelling. Below we provide S-plus and SAS codes for random effects (RE) and mixed effects (A, B, and C) modelling for all four models. The data set **growth** contains an identifier for each subject (**id**), age at menarche (**age.at.menarche**), two variables for time relative to event (**time.before** and **time.after**), a variable indicating parental obesity (**parental.obesity**), and an outcome variable—%body fat (**y**).

**S-plus CODE**

```
function (MODEL)
{
  if (MODEL == "RE") {
    ## random effects model: y ~ time.before + time.after
    m <- lme(fixed = y ~ time.before + time.after, random
      = ~ time.before + time.after, cluster = ~ id, data
      = growth, na.action = na.omit)
  }
  if (MODEL == "A") {
    ##
    ## Model A: y ~ time.before + time.after + time.after^2
    m <- lme(fixed = y ~ time.before + time.after
      + I(time.after ^2), random = ~ time.before + time.after
      + I(time.after^2), cluster = ~ id, data = growth, na.action
      = na.omit)
  }
  if(MODEL == "B") {
    ##
    ## Model B: y~ time.before + time.after + parental.obesity
    m <- lme(fixed = y ~ time.before + time.after
      + parental.obesity, random = ~ time.before + time.after,
      cluster = ~ id, data = growth, na.action = na.omit)
  }
  if (MODEL == "C") {
    ##
    ## Model C: y~time.before + time.after * parental.obesity
    m <- lme(fixed = y ~ time.before + time.after *
      parental.obesity, random = ~ time.before + time.after,
      cluster = ~ id, data = growth, na.action = na.omit)
  }
  print(m)
  print(anova.lme(m))
}
```

**SAS CODE**

```
data growth;
set save.fatrw;
run;

Model RE
proc mixed data = growth;
class id;
model y = time.before time.after / s;
random intercept time.before time.after / type = un subject =
id;
run;

Model A
proc mixed data = growth;
class id;
model y = time.before time.after time.after*time.after / s;
random intercept time.before time.after time.after*time.after /
type = un subject = id;
run;

Model B
proc mixed data = growth;
class id parental.obesity;
model y = time.before time.after parental.obesity / s;
random intercept time.before time.after / type = un subject
= id;
run;

Model C
proc mixed data = growth;
class id parental.obesity;
model y = time.before time.after parental.obesity
time.after*parental.obesity / s;
random intercept time.before time.after time.after / type
= un subject = id;
run;
```

## Commentary: Are piecewise mixed effects models useful in epidemiology?

Rebecca Hardy

The call for applied researchers to learn how to use mixed effects models in the paper by Naumova, Must and Laird<sup>1</sup> in this issue of the *International Journal of Epidemiology* is to be welcomed. The value of longitudinal studies, in which measurements on the same sample of individuals are taken repeatedly over time,

is well understood. Such studies are costly, both financially and in terms of the time and effort required on behalf of study members and researchers. Correct analysis of the often vast amount of data collected is therefore vital. Despite considerable discussion regarding the analysis of repeated measures in the statistics literature, the use of more complex models in medical and epidemiological journals remains scarce. As well as researchers being encouraged to use more complex models, journal editors should also welcome their use where appropriate.

MRC National Survey of Health and Development, Department of Epidemiology and Public Health, Royal Free & University College Medical School, 1-19 Torrington Place, London WC1E 6BT, UK. E-mail: rebecca.hardy@ucl.ac.uk

Mixed effects models can be used to analyse straightforward repeated outcome measure data over time, many examples of which exist in epidemiology. However, the specific mixed effects model presented by Naumova *et al.*<sup>1</sup> is more interesting. The idea of using a time scale other than age or calendar time is introduced, as is the use of a piecewise model. The piecewise model allows separate slopes to be fitted to the observations representing the periods before and after a 'critical period' or 'event'. The extension of the basic model to determine the individual characteristics that relate to differences in the increase in fat after menarche is particularly valuable to the understanding of obesity development. Understanding how early life and childhood growth influence not only obesity, but also later disease development has been highlighted by the recent interest in life course epidemiology.<sup>2</sup> For example, early age at menarche<sup>3</sup> and tall adult height<sup>4</sup> have been consistently related to increased breast cancer risk, but the mechanism underlying the associations may be due to childhood growth. Women who were both big babies and tall at 7 years have been found to have a greater risk of premenopausal breast cancer than others.<sup>5</sup> Hence, understanding how growth patterns, in relation to menarche, are influenced by factors such as childhood diet and socioeconomic status is important in breast cancer aetiology. Menopause is another obvious event to which the proposed models could be applied. Insight into the debates regarding the influence of menopause on cardiovascular risk factors<sup>6</sup> and on psychological health<sup>7</sup> could be gained. The 'event' could even be social such as loss of job or divorce, both of which may influence health.

The additional information obtained from mixed effects models regarding between-individual variability can be of interest in its own right.<sup>8</sup> Assuming that the individual post menarcheal slopes are normally distributed, approximately 95% of girls' slopes will lie within  $\pm 1.96$  standard deviations of the overall average slope. For the example presented by Naumova *et al.*<sup>1</sup>, the limits would be 0.62–4.30% and hence approximately 2.5% of girls have an increase in fat above 4.3% per year. The percentage of girls above or below any given slope can be obtained by reference to the standard normal distribution. It is this idea that is utilized when Naumova *et al.*<sup>1</sup> observe that there are virtually no girls with a negative slope after menarche but 37% with a negative slope before menarche. Such information could be useful when planning obesity prevention strategy. The correlation between body fat at menarche and post menarche slope of  $-0.59$  suggests that the greater the fat at menarche the slower the increase in fat afterwards.

It may be unclear to the non-statistician why the desired statistical property of *efficiency* is important. In practice, the confidence intervals of the estimates in an efficient analysis will be narrower than those from a less efficient one and hence use of the more efficient analysis means there is greater power to detect a given effect. Efficiency can therefore also have implications for study design and sample size requirements. Although some improvement in efficiency is achieved simply through the model fitting procedure, additional gains will often be made simply through mixed effects models allowing more individuals to be included in the analysis than standard methods. This is due to their ability to handle missing outcome data and measures collected at different time points for different individuals. This represents a major practical advantage as unbalanced data is unavoidable in longitudinal epidemiological studies. Although,

in technical terms, such models can always handle missing data, assumptions about the missing data mechanism must be made in order for the inferences to be valid. Specifically, the outcome data must be missing at random which means that the probability of a measure being missing may depend upon observed, but not unobserved, measures. Although impossible to test, the feasibility of such an assumption should be considered in each analysis as it may be more realistic in some cases than others. In studies of cognitive function in the elderly, a missing cognitive test score may be more likely if function has deteriorated. It should also be noted that the mixed effects models cope far less easily with missing covariate information.

With the increased ease of implementation, further aided by the SAS programs provided in the paper,<sup>1</sup> wider use of such models should be encouraged. It has to be acknowledged, however, that these models are complex and this is perhaps highlighted by the fact that Naumova *et al.*<sup>1</sup> still require a certain amount of mathematical notation in their explanation. Effort is therefore required on behalf of the applied researcher wishing to utilize such models correctly. The use of graphical displays at all stages of the analysis can certainly aid understanding. The figures presented by Naumova *et al.*<sup>1</sup> highlight the two components of the mixed effects model, the subject-specific growth curves and the population averages, which are key to such models. Finally, as with all statistical techniques, but perhaps particularly with one as flexible as the mixed effects model, caution is required and all analyses should be carried out with due thought to the specific hypothesis under consideration. The added complexity of having to specify correctly the random effects component of the model as well as the fixed effects means that model checking in addition to that usually carried out is required.<sup>9,10</sup>

## References

- 1 Naumova EN, Must A, Laird NM. Evaluating the impact of 'critical periods' in longitudinal studies of growth using piecewise mixed effects models. *Int J Epidemiol* 2001;**30**:1332–41.
- 2 Kuh D, Ben-Shlomo Y (eds). *A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-health from Early to Adult Life*. Oxford: Oxford University Press, 1997.
- 3 Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiol Rev* 1993;**15**:36–47.
- 4 Wang DY, De Stavola BL, Allen DS *et al.* Breast cancer is positively associated with height. *Breast Cancer Res Treatment* 1997;**43**:123–28.
- 5 DeStavola BL, Hardy R, Kuh D, dos Santos Silva I, Wadsworth M, Swerdlow AJ. Birthweight, childhood growth and risk of breast cancer in a British cohort. *Br J Cancer* 2000;**83**:964–68.
- 6 Matthews KA, Kuller LH, Sutton-Tyrrell K, Change Y-F. Changes in cardiovascular risk factors during the perimenopause and postmenopause and carotid artery atherosclerosis in healthy women. *Stroke* 2001;**32**:1104–11.
- 7 Klein P, Versi E, Herzog A. Mood and menopause. *Br J Obst Gynaecol* 1999;**106**:1–4.
- 8 Hardy R, Wadsworth M, Kuh D. The influence of childhood weight and socioeconomic status on change in adult body mass index in a British national cohort. *Int J Obesity* 2000;**24**:725–34.
- 9 Beacon H, Thompson S. Multi-level models for repeated measurement data: application to quality of life data in clinical trials. *Stat Med*; **15**:2717–32.
- 10 Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Stat Med* 1997;**16**:2349–80.