Chapter 32

Multilevel growth curve analysis for quantitative outcomes

Douglas A. Luke

Multilevel growth curve modeling is a powerful and flexible statistical technique which can be used to model longitudinal data. The primary purposes of growth curve modeling are to describe the form and structure of change in a quantitative dependent variable over time, and to explore the interindividual and intraindividual predictors of this change. Growth curve modeling is a type of multilevel modeling, based on a mixed effects statistical model, which treats multiple observations as nested within individuals. Growth curve modeling has numerous statistical advantages for analyzing longitudinal data. In particular, it can handle missing data and longitudinal designs where observations occur at different time points across individuals. Growth curve models can be fit with any statistical software that includes mixed effects or multilevel modeling procedures. Multilevel growth curve modeling is one of the most powerful and flexible ways to analyze longitudinal data.

1 Introduction

Growth curve modeling is a flexible and powerful way to analyze longitudinal data. The term "multilevel growth curve" recognizes the fact that growth curve modeling is a type of multilevel model where observations are nested within individual cases. The use of the term growth curve arises out of psychology, where this type of multilevel modeling was first used to describe developmental growth of a variety of psychological characteristics (Bryk and Raudenbush, 1987; McArdle and Nesselroade, 2002). However, as we shall see, growth curve modeling can be applied to any form of longitudinal data where the interest is in change—no formal conception of a "growth" process is required.

Multilevel growth curve modeling uses a mixed effects general linear modeling approach to estimate the statistical model. Growth curve models can also be estimated using a latent construct approach via structural equation modeling (SEM) software. This SEM approach is not covered in this chapter. Interested readers can see the excellent introduction to latent growth curve models by Terry Duncan and his colleagues (1999). See also Stoolmiller, Chapter 31, in this volume.

Multilevel growth curve analysis has a number of important strengths. First, it allows flexible statistical modeling that can more closely match the underlying longitudinal theoretical framework. In particular, multilevel growth curve modeling can help disentangle questions about interindividual predictors (e.g., Do children who go to pre-school show quicker

. Handbook of Longitudinal Research : Design, Measurement, and Analysis Across the Social Sciences.
Burlington, MA, USA: Academic Press, 2007. p 558
http://site.ebrary.com/lib/wustl/Doc?id=10206036&ppg=558

mastery of reading skills in elementary school compared to children who do not go to preschool?) from intraindividual predictors (e.g., Does receipt of positive feedback speed up the acquisition of particular reading skills for individual children?).

The primary purpose of this chapter is to introduce growth curve modeling techniques in an applied way so that the interested reader can see their potential for longitudinal data analysis. In the next section some basic considerations about research design and data management in growth curve modeling are discussed. Following this, longitudinal data from the National Longitudinal Survey of Youth are used to illustrate the basic steps in building and evaluating a growth curve model. Finally, the chapter concludes with a short appendix covering software that can be used to fit growth curve models.

2 Research design and data management

Growth curve models are based on longitudinal data from longitudinal research designs. Longitudinal data are made up of observations of one or more dependent and independent variables which are measured on the same individuals at multiple points in time. Technically, pre-post data that are obtained at two different time points are longitudinal. However, no real longitudinal research questions can be addressed with such data. Questions about the form of change over time, in particular, can only be answered with data that are measured at three or more time points (Singer and Willett, 2003).

Longitudinal data may be obtained from either experimental or observational studies. For example, a clinical trial study of the effects of an educational campaign designed to promote screening for prostate cancer would collect longitudinal data from participants over time after enrolling in the study. The primary hypothesis would be that participants receiving the new educational materials would show higher rates of screening over time than participants in a control condition. Longitudinal observational studies are also extremely common. For example, using health surveys of adolescents, investigators could track substance use patterns over time. One longitudinal hypothesis could be that students who transfer schools may show steeper increases in substance use over time than students who remain in the same school. From the perspective of the statistical analyst, there is no difference between experimental and observational longitudinal data. The primary difference is in the interpretation of the results—e.g., much stronger claims for causality may be made for experimental longitudinal data than from observational data.

2.1 Data management

Data management for growth curve and other types of longitudinal data analysis can become somewhat complicated. However, all longitudinal datasets will have certain core features. First, longitudinal datasets will have five basic types of variables: an ID variable, one or more longitudinal dependent variables, one or more variables containing time information, time-varying predictors, and time invariant predictors. A dataset used for growth modeling will always have at least the first two types, but the presence of the different types of predictor variables will depend on the study design and research questions.

Although longitudinal data are often initially collected and stored in different data files, eventually the data will be brought together for analysis. There are two common formats for storing longitudinal data, illustrated in Table 32.1. In the "wide" data format, each record in the database is a separate individual. Multiple observations on the same individual are stored in different variables (e.g., weight1, weight2, etc.) in the same case. However, most multilevel software packages will expect to see longitudinal data in a different format, where data

are stored in one observation per record. In this "tall" format each observation gets its own record, and any longitudinal data are stored in one variable (e.g., weight). Most general-purpose statistical packages provide routines that can relatively easily restructure the data from one format to the other. Notice that in the observation record format, the multilevel structure of the longitudinal data is apparent: multiple observations are nested within individuals (see below).

2.2 Introduction to the NLSY97 dataset

The data used to provide examples for this chapter are taken from the National Longitudinal Survey of Youth 1997 Cohort (NLSY97). The NLSY97 is part of a series of surveys funded by the US Bureau of Labor Statistics and designed to gather longitudinal data on the labor market experiences of US youth and adults. The NLSY97 examines the transition from school to work for a nationally representative sample of youth who were born from 1980 to 1984. The youths were ages 12 to 17 during the first wave of data collection. 8984 participants were interviewed in 1997, and annual interviews were conducted for the next seven years. The sample size for round 7 was 7756, and the overall retention rate was 86.3%. The NLSY97 collected information on a wide variety of educational, work, and health areas. With the large sample size, number of variables, and up to seven time points for each participant, the NLSY97 is an ideal data source for exploring growth curve modeling. See http://www.bls.gov/nls for more information.

For this chapter, data were extracted and downloaded from the complete seven-year NLSY97 public dataset. We will be focusing on developing growth models for two dependent variables: BMI and Total Substance Use Days. BMI is the body mass index and is an important risk factor for a wide variety of health conditions related to obesity. BMI was not measured directly in the NLSY97, but is based on

Table 32.1 Comparison of the "individual record" (wide) and "observation record" (tall) data structures

Individual record structure							
ID	*Gender*	*Age1*	*Weight1*	*Age2*	*Weight2*	*Age3*	*Weight3*
001	M	12	125	13	129	14	137
002	F	12	101	13	103	14	108

Observation record structure				
ID	*Time*	*Gender*	*Age*	*Weight*
001	1	M	12	125
001	2	M	13	129
001	3	M	14	137
002	1	F	12	101
002	2	F	13	103
002	3	F	14	108

self-reported measures of height and weight. The formula for BMI is

$$BMI = \left(\frac{Weight\ in\ pounds}{(Height\ in\ inches) \times (Height\ in\ inches)} \right) \times 703$$

The NLSY97 asked youth to report the number of days in the past 30 days that they had used alcohol, marijuana, or smoked cigarettes. We combined these three measures to form a total substance use risk variable, called Substance Use Days, that can range from 0 to 90. The higher the number, the more often the youth is reporting using substances in the past month. We will use hierarchical linear modeling to examine how each of these variables change over time as youths age, and we will also explore how certain covariates predict interindividual differences in change patterns over time. The covariates include individual characteristics such as gender and race, as well as one important time-varying predictor, transition to a new school.

3 Building the multilevel growth model

3.1 Framing a growth curve model as a multilevel model

In a traditional regression model, variability of the dependent variable is accounted for either by the predictor variables, or else put into an undifferentiated individual error term. In a multilevel statistical model, as the name suggests, we are able to partition variability across multiple levels. So, for example, if we want to understand reading achievement by students in multiple classrooms, using a multilevel model we can account for variability that exists between students (level 1) and also variability between classrooms (level 2). That is, students are nested in classrooms, and we can build statistical models that reflect that reality.

Growth curve models are simply a special type of multilevel model. Here, multiple observations across time are nested within individuals. As we stated earlier, a principal advantage of multilevel modeling is its ability to account for nonindependence of observations due to nesting. So, just as we might expect students in the same classroom to be more similar to one another than would be expected by chance (thus violating the traditional independence assumption), we certainly expect multiple observations of the same person to be more alike. Growth curve modeling using hierarchical linear models can appropriately account for this nonindependence of observations across time.

The following system of equations shows a basic growth curve model as a multilevel model:

$$Y_{ti} = \beta_{0i} + \beta_{1i} T_{ti} + \varepsilon_{ti}$$
$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$

Here we are modeling some dependent variable Y, measured at time t on individual i. The first line of the model can be considered as the level 1 portion of the model, and looks similar to a typical multiple regression model. The only level 1 predictor included in this basic growth curve model is T, which is the time variable. For this reason this model is sometimes called an *unconditional linear growth curve model*. It is unconditional in that there are no predictors, other than the time variable. It is linear in that β_{1i} only captures the linear relationship between time and the dependent variable.

The most important difference between the above model and a traditional multiple regression model can be seen by considering the presence of the i subscripts in the level 1 portion of the model. Both the intercept and slope betas have i subscripts, indicating that we are allowing the intercepts and slopes to vary across individuals. That is, each individual in the dataset

is allowed to have his or her own growth curve! For this reason, we often call the level 1 part of the growth curve model the *intraindividual* part of the model.

In a multilevel model, the parameters in the first level of the model become outcomes in the second level of the model. In the unconditional linear growth model, the intercept for a particular individual (β_{0i}) is predicted by the grand mean of all the individual intercepts (γ_{00}) plus the variability of the individual intercepts around the grand mean (u_{0i}). Similarly, the slope for a particular individual (β_{1i}) is predicted by the grand mean of all the individual slopes (γ_{10}) plus the variability of the individual slopes around the grand mean (u_{1i}). The level 2 part of the model is called the *interindividual* part of the model, because it can be used to model predictors of change between individuals.

Instead of using a system of equations to specify the multilevel model, we can substitute the level 2 parts of the model into the level 1 equation. After substituting and rearranging the terms, we get the following:

$$Y_{ti} = \underbrace{[\gamma_{00} + \gamma_{10} T_{ti}]}_{fixed} + \underbrace{[u_{0i} + u_{1i} T_{ti} + \varepsilon_{ti}]}_{random}$$

This single prediction equation form of the multilevel model is called the *mixed effects model*, because it shows how the model is based on both fixed effects (the gammas γ) and random effects (the variance components ε and u). Although it is harder to discern the multilevel structure of the model when it is in this form, it more clearly states what components are actually being modeled. This form of the model also closely corresponds to the output of the various multilevel modeling software packages. It is advantageous to be able to construct and interpret multilevel models using both types of equations. Fortunately, hierarchical linear modeling software such as HLM allows you to see the models in both forms.

The unconditional linear growth model, presented above, is only one of an innumerable set of possible growth curve models. This simple model can be extended by adding predictors at either of the levels of the model, as well as by making decisions about which random effects to include in the model. It is difficult at this point to know how to make these decisions. So in the next two sections we will look at how to extend the model by considering two fundamental questions about growth curve models. First, how can we describe the form of change over time? Second, what factors influence intra- and interindividual patterns of change?

3.2 Describing the form of change

A starting point for most growth curve models is to describe the form or shape of change in the dependent variable of interest. The purpose of this first step may simply be descriptive, or it might be to address a specific scientific question (e.g., "Does the increase in BMI during adolescence follow a quadratic form?").

Before jumping into fitting and testing specific multilevel growth models, it is advisable to spend some time thinking theoretically about expected patterns of change. This can help guide the often complicated process of model selection. In addition, it is always a good idea to examine the data to see what the individual raw growth curves look like. Figure 32.1 presents 40 randomly chosen plots of the raw growth curves of BMI from the NLSY97 data. It is apparent that BMI levels vary substantially between youth. However, it appears that for many of the youth, BMI tends to go up as they get older. Although the patterns are not consistent across youth, it appears that changes in BMI may not proceed in a simple linear fashion. So we might want to examine more complicated growth models that describe nonlinear change.

Figure 32.1 also reveals that a number of the participants do not have measurements for all seven time points. In fact, just in this random sample we see one person with only one measurement, and a couple of people with

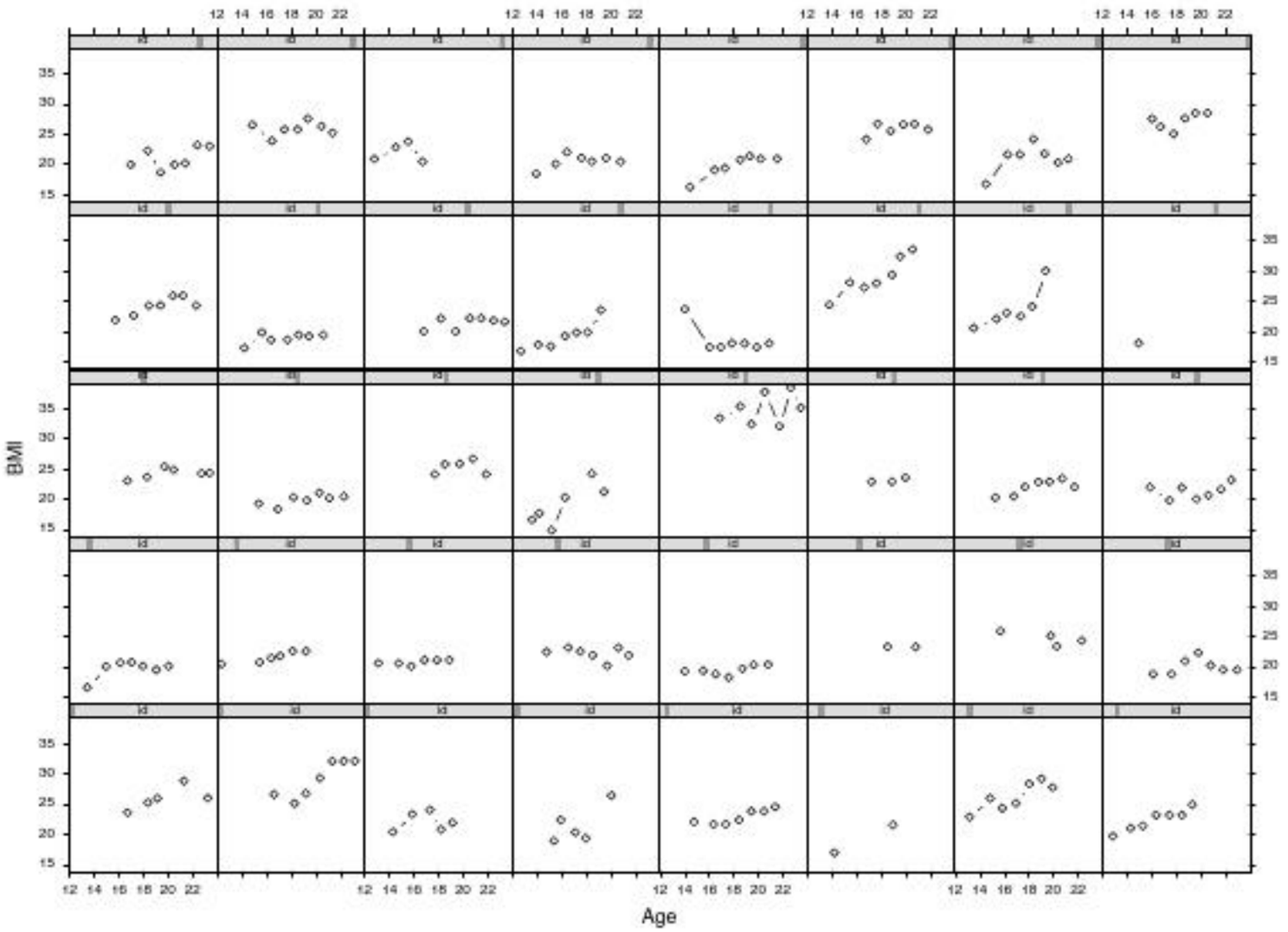

Figure 32.1 Individual growth curves of BMI by age for 40 random cases

only two measurements. Very close examination of the figure also reveals that the period between measurements varies across individuals. That is, although youth are interviewed *on average* once a year in the NLSY97, the actual time between interviews can be quite a bit shorter or longer for particular individuals. These two common aspects of real world longitudinal datasets, varying number of measurements and varying time between measurements, pose severe or even fatal challenges for traditional longitudinal statistical approaches such as repeated measures ANOVA. However, multilevel modeling can handle this type of "messy" data without any problem. This is, in fact, one of the primary reasons that multilevel modeling is now a preferred analytic approach for growth curve models.

Unless you have a very specific hypothesis about the form of the change, a reasonable approach to model building is to start simple, and then build more complex models. For this reason, we start by fitting the above unconditional linear growth model to BMI:

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(Age12)_{ti} + \varepsilon_{ti}$$

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

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

In this model we are predicting BMI scores for individuals across time. Time is measured using the age of the youth at the time of the interview after subtracting 12. (We will explain the reason for this below.) Both the intercept and time slope are allowed to vary across individuals. With this model we will be able to see how much BMI goes up or down as youths age. Given our examination of the raw growth curves, we might expect the time slope to be positive.

The results of this model are presented under Model 1 in Table 32.2. This table summarizes much of the useful information from a growth model. The top of the table presents estimates of the fixed effects. The coefficients are estimates of the two gammas. The average intercept across all individuals is 20.63. This means that the expected BMI for any individual when the age variable is 0 is approximately 21. This helps explain why we subtracted 12 from the age at interview variable. If we used the raw age at interview, then the estimate of the intercept would be interpreted as the expected BMI score when a person was age 0 (i.e., a newborn). This, of course, is not a useful or interpretable estimate. By subtracting 12 from each age at interview score, we get a new interpretation of the intercept—the expected BMI value at age 12. We picked age 12 because this is approximately the youngest age for which there are data in the NLSY97 dataset. This is an example of *centering* a predictor variable. There has been a lot written about centering variables in multilevel models (see, e.g., Paccagnella, 2006). Although the topic can get quite complicated, the most important reason to center predictor variables is to produce fixed effects estimates that are more interpretable than would otherwise be the case.

The linear fixed effect estimate of 0.54 tells us that for each additional year of age, we would expect BMI to increase by about half a point for each individual. This suggests that during the teen and young adult years, youth are becoming more overweight as they age. Along with the coefficient estimates, their associated standard errors, *t*-tests, and *p*-values are presented. Hypothesis testing of the individual fixed effects parameters can thus be done using traditional methods.

The middle rows of Table 32.2 present the random effects part of the model. These are presented in the form of variance components, and can be thought of as unmodeled variability. The variance component of 16.64 tells us that there is a large amount of variability of individual BMI scores around the average starting point of 20.63. This confirms what we saw in Figure 32.1, where we saw some people with quite low BMI scores, and others with high BMI scores. The much smaller linear variance component of 0.23 suggests that there is much less variability of the slopes across individuals. One way to view this is that there is much more variability left over to model (with predictor variables) of the *level* of BMI, than there is of the *slope* of BMI on age. Finally, the level 1 variance component estimate of 3.31 suggests that there is a moderate amount of intraindividual variability. This suggests that the individual observations may be bouncing around the linear regression line. This could be due to instability of the BMI measurements. Another possibility is simply that the simple linear growth model is not a good fit with the data.

In addition to the variance component estimates, some multilevel modeling software packages will produce statistical tests of these components. Here we see the chi-square tests and associated *p*-values produced by HLM. However, these statistical tests should be viewed with caution. First, variance components are bounded at 0, so their distributions are not normal. Second, it is not clear exactly what the meaning of a significant variance component should be—after all, we generally expect variances to be nonzero. Rather than focusing on the *p*-values of the variance components, it is usually more fruitful to interpret their

Table 32.2 Three growth models for change of BMI

Fixed effects	*Model 1 – Linear*				*Model 2 – Quadratic*				*Model 3 – Cubic*			
	Coef.	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*
Intercept (γ_{00})	20.63	0.050	413.7	0.000	20.172	0.067	302.7	0.000	19.77	0.088	225.5	0.000
Linear (γ_{10})	0.54	0.006	84.6	0.000	0.722	0.020	36.2	0.000	1.00	0.046	21.4	0.000
Quadratic (γ_{20})					−0.016	0.002	−9.7	0.000	−0.07	0.008	−8.2	0.000
Cubic (γ_{30})									0.003	0.000	6.4	0.000
Random effects	*Variance component*		χ^2	*p*	*Variance component*		χ^2	*p*	*Variance component*		χ^2	*p*
Intercept (u_{0i})	16.64		39260	0.000	16.54		14437	0.000	16.46		14381	0.000
Linear (u_{1i})	0.23		26028	0.000	0.93		10439	0.000	0.91		10384	0.000
Quadratic (u_{2i})					0.006		10403	0.000	0.006		10360	0.000
Cubic (u_{3i})									—[a]		—	—
Level 1 (ε_{ti})	3.31				3.10				3.10			
Model fit												
Deviance	260895.1				260257.6				260209.0			
Parameters	6				10				11			
AIC	260907.1				260277.7				260231.0			
BIC	260960.6				260366.7				260329.0			

[a]Cubic effect set to fixed to avoid convergence problems.

sizes rather than their significance (Pinheiro and Bates, 2000).

3.3 Assessing nonlinear change patterns

Model 1 tells us that BMI increases with age, but there is still a lot of variability both across and within individuals. Figure 32.1 suggested that increases in BMI may not be strictly linear in form, so the next step is to build models that will assess the extent to which the form of the change of BMI is nonlinear. There are a number of ways of building such curvilinear models. One of the simpler approaches is to build a polynomial growth model by adding quadratic, cubic, quartic terms, and so on, to the base linear model. For a dataset with k time points, in principle k-1 polynomial terms can be fit. However, in practice growth models are rarely built that go beyond cubic or quartic components. First, in most areas of the social and health sciences theories are not rich enough to suggest or explain such high-level polynomial models. Second, in many real-world datasets quadratic or cubic models explain most of the intraindividual variability, and it is unusual to have underlying variability that requires more complicated models.

To fit polynomial models, you simply add the appropriate time variable raised to the degree of the polynomial. So, a quadratic model would include Time, and Time-squared. A cubic model would include Time, Time-squared, and Time-cubed, and so on. The following equation is for a quadratic polynomial model of change of BMI:

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(Age12)_{ti} + \beta_{2i}(Age12)^2_{ti} + \varepsilon_{ti}$$

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

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

$$\beta_{2i} = \gamma_{20} + u_{2i}$$

Models 2 and 3 listed in Table 32.2 present the results of fitting a quadratic and cubic model, respectively. For both models, the individual coefficients are highly significant, suggesting that a curvilinear model is more appropriate than a simple linear model. In polynomial models, the meaning of the coefficients changes. For example, in a quadratic model, the linear coefficient for time no longer represents a constant change rate. Instead, it now represents the instantaneous rate of change at the point that time = 0. The quadratic coefficient tells us how fast the instantaneous rate of change itself changes. This can be thought of as a *curvature* parameter (Singer and Willett, 2003). Instead of interpreting each coefficient individually in a polynomial model, it is often more informative to plot the prediction curves for the model based on the fitted coefficients. Figure 32.2 presents the prediction curves for BMI for the three models presented in Table 32.2.

The figure shows that BMI increases steadily as youths age. The curvilinear models both suggest that BMI rises faster at early ages, from about 12 to 15. The quadratic model suggests that after about the age of 22 the increase in BMI starts slowing down. To see how these models fit the data at the two age extremes, individual marks were added to the plot that represent the raw average BMI scores for that age group. These marks suggest that the quadratic and cubic models fit the data pretty well for the

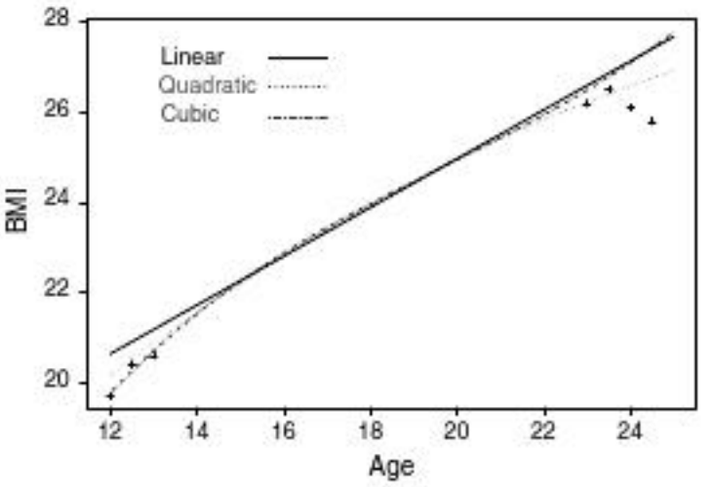

Figure 32.2 BMI prediction curves for linear, quadratic, and cubic growth curve models

Table 32.3 Three growth models for Substance Use Days

Fixed effects	*Model 1 – Linear*				*Model 2 – Quadratic*				*Model 3 – Cubic*			
	Coef.	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*
Intercept (γ_{00})	0.030	0.161	0.2	0.852	−2.846	0.174	−16.3	0.000	−0.145	0.214	−0.7	0.497
Linear (γ_{10})	1.655	0.030	55.4	0.000	2.777	0.083	33.44	0.000	0.677	0.173	3.9	0.000
Quadratic (γ_{20})					−0.092	0.007	−13.1	0.000	0.326	0.035	9.3	0.000
Cubic (γ_{30})									−0.024	0.002	−11.9	0.000
Random effects	*Variance component*		χ^2	*p*	*Variance component*		χ^2	*p*	*Variance component*		χ^2	*p*
Intercept (u_{0i})	80.75		15080	0.000	27.25		6486	>0.500	27.11		6426	>0.500
Linear (u_{1i})	4.26		21205	0.000	26.51		9218	0.000	26.35		9147	0.000
Quadratic (u_{2i})					0.162		9579	0.000	0.161		9508	0.000
Cubic (u_{3i})									—[a]		—	—
Level 1 (ε_{ti})	100.24				90.02				89.74			
Model fit												
Deviance	431258.6				428258.7				428116.0			
Parameters	6				10				11			
AIC	431270.6				428278.7				428138.0			
BIC	421324.0				428367.8				428236.0			

[a]Cubic effect set to fixed to avoid convergence problems.

early ages, while the quadratic model does a better job for young adults. (The means for the youngest, age 12, and oldest, age 24.5, groups are based on very small numbers of cases, so they should be interpreted with caution.) An interesting thing to note about these models is that from about the ages of 15 to 22, all three models would lead to virtually the same predicted values.

Table 32.3 and Figure 32.3 present the results for the same set of growth curve models applied to the Substance Use Days dependent variable. As youths age, we see that the number of substance use days also goes up. Again, we find that there are significant curvilinear components to the change over time. Figure 32.3 shows that a linear model does a particularly poor job of predicting substance use for the oldest members of the sample. Conversely, the quadratic model gives impossible predictions for kids aged 12 or 13. The cubic model may do the best job, and it describes a type of S-curve that seems reasonable for this dependent variable. When kids are very young, substance use is near zero and changes slowly. During the teen years substance use increases significantly, but as adulthood approaches, substance use appears to level off.

3.4 Model diagnostics, fit and selection

In addition to examining individual parameters and their associated *p*-values, it is usual to examine model diagnostics to see how well the fitted model matches its underlying assumptions, and then to examine various fit indices to see how well the overall model fits the data.

Diagnostics for growth curve models of quantitative dependent variables are very similar to those examined for multilevel modeling. Two common assumptions that can be easily checked are normality of errors (residuals) and homoscedasticity. Figure 32.4 shows two diagnostic plots on a random 2% sample (~1000 cases) of the quadratic BMI growth model (Model 2 from Table 32.2). The Q-Q plot on the left side tells us that although the residuals are symmetric, they are more kurtotic (higher central peak, smaller tails) than we would expect with independent and normally distributed errors. This suggests that our model may require a more complex covariance structure than we assumed. (Our model was fit assuming a compound symmetry covariance matrix. For details on how to fit growth curve models with other covariance structures, see Singer and Willett, 2003). On the right side we plot the residuals against the fitted (predicted) BMI values. This plot shows no evidence of a fan shape, and strongly suggests that this model does not have problems with heteroscedasticity. Luke (2004) provides more examples of how to use graphical exploration of residuals to examine the assumptions of multilevel models.

Growth curve models for quantitative dependent variables are typically fitted using some form of maximum-likelihood estimation (Laird, 1978). Simply stated, this type of estimation works by maximizing a likelihood function that assesses the joint probability of simultaneously observing all of the sample data, assuming a certain set of fixed and random effects. An important product of the estimation process is a number obtained by multiplying the natural log of the likelihood by −2. This number, sometimes called the *deviance* or designated as

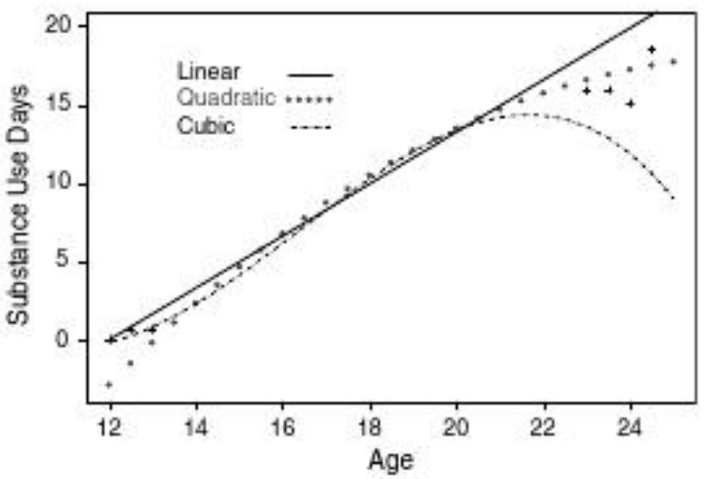

Figure 32.3 Substance Use Days prediction curves for linear, quadratic, and cubic growth curve models

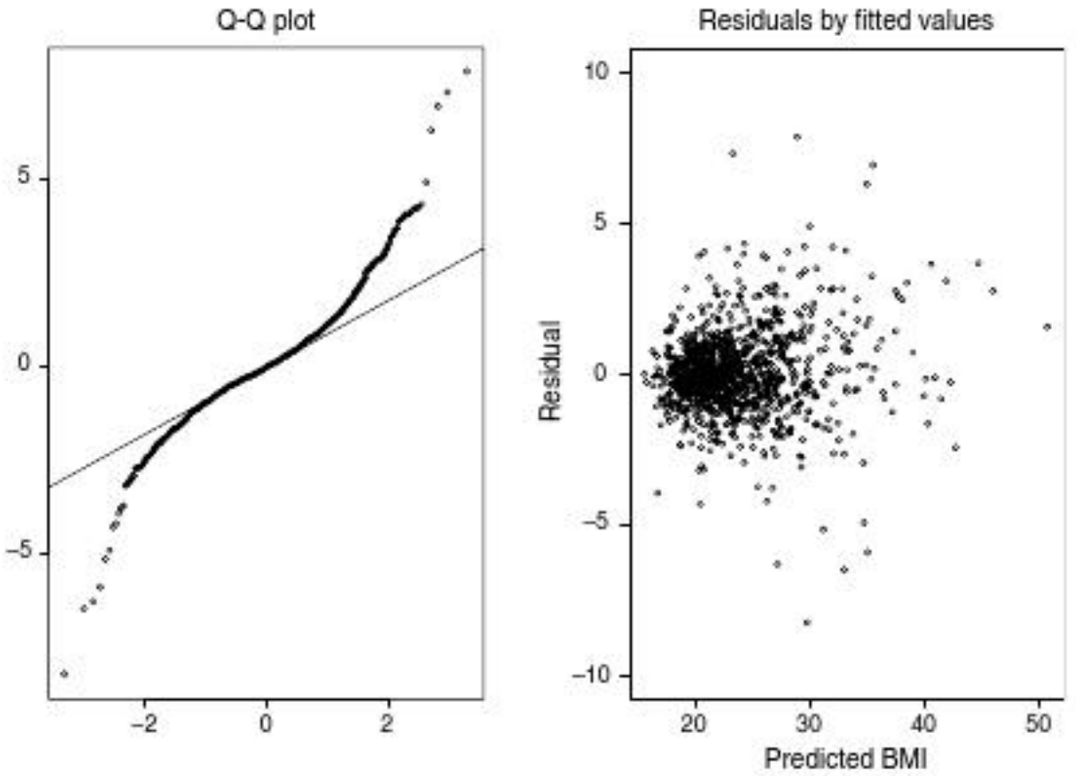

Figure 32.4 Two diagnostic plots for the quadratic BMI growth curve model

−2LL, is a measure of the discrepancy between the observed data and the fitted model. The deviance for any one model cannot be interpreted directly, but it can be used to compare multiple models to one another.

The model comparison can be done between two models fit to the same data, where one of the models is a subset (has fewer parameters) of the other. The difference of the deviances from each model is distributed as a chi-square statistic with degrees of freedom equal to the difference in the number of parameters estimated in each model. For example, we can compare Model 3 to Model 1 for BMI (Table 32.2) to see if the nonlinear change model is better than the simpler linear change model. The difference between the two deviances is 886.1 (260895.1 − 260209.0); this value is highly significant with $df = 5(11-6)$. This tells us that the more complicated model is a significantly better fit to the data.

One disadvantage of the deviance (−2LL) is that a model fit to the same data with more parameters will always have smaller deviance. This is generally good, because smaller deviance implies a better fit to the data. However, we can always get better fit by adding more predictors. We also want to choose the simplest model that describes the data; i.e., the model with the fewest parameters. Two widely used fit indices have been developed that are based on the deviance, but incorporate penalties for a greater number of parameters: The Akaike Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC) (Akaike, 1987; Schwarz, 1978). For both of these indexes, smaller is better. Also, an important advantage of these two criteria is that they can be used to compare two models fit to the same dataset, even if one is not a subset of the other. The AIC and BIC are listed in Tables 32.2 and 32.3 for our change models. In both cases

these criteria indicate that the cubic models are better models than the simpler linear models. For more details on how to calculate and use AIC and BIC, see Luke (2004).

3.5 Different choices for scaling time

The concept of time is critical for growth curve models. Therefore, it is also critical to think carefully about how to operationalize, measure, and model time in a growth curve model. Time can be defined many ways for any particular study—and these different conceptions may not all equally represent the theory or research question under consideration, and different operationalizations of time may lead to different model results. For example, consider Table 32.4, which shows four different ways that time may be assessed for the NLSY97 study: the interview wave (from 1 to 7), the actual age at the time of the interview, the age after subtracting 12, and the grade that the student is in. If the investigator is interested in the underlying physiological and cognitive changes that influence body weight, then Age or Age12 might be appropriate conceptions of time. On the other hand, if one wants to understand how the changes in school environment may influence drug use, then perhaps Grade would be a useful operationalization of time. However, it is hard to think of a research question that would be usefully served by using Interview Wave as a measure of time. This is an arbitrary time measurement that is based on the logistics of the study, not a physical or social reality.

Table 32.4 Examples of different definitions of time

ID	*Interview wave*	*Age*	*Age12*	*Grade*
001	1	13	1	7
001	2	15	3	8
001	3	16	4	8
001	4	16	4	9
001	5	18	6	11
001	6	19	7	13
001	7	20	8	14
002	1	15	3	10
002	2	16	4	11
002	3	16	4	12
002	4	18	6	13
002	5	19	7	14
002	6	20	8	15
002	7	21	9	15

Table 32.5 presents the fixed effects results of a linear BMI growth model for three different definitions of time. The results for Model 3 on the right side of the table are for the Age12 variable, and are the same as displayed in Table 32.2. Model 1 presents a growth model where Interview Wave is used for time. The results are similar to Age12, but not identical. They are similar in that on average across all subjects each interview wave is approximately one year apart. Both models show that BMI increases about half a point a year. However, consideration of the AIC and BIC scores shows that the model with Age12 is doing a better job of describing the observed data. This is not surprising, because Age12 provides information not just on the order of the interviews, but also reflects an accurate measure of the actual amount of time that has passed between each interview for each participant.

The difference between Models 2 and 3 is simpler and more subtle. Age is the raw age, while Age12 is raw age minus 12. Subtracting (or adding) a constant from a predictor variable is a way of *centering* the predictor variable. Centering is typically done in one of three ways: (1) by subtracting a meaningful constant, as we have done here with Age12; (2) by subtracting a grand mean; or (3) by subtracting a group mean. This third type of centering is more complicated than the other two, but is relatively uncommon for growth models (where the group is each individual). In growth modeling, centering is typically done by subtracting a constant or grand mean, and this has two advantages. First, centering is typically done so that the interpretation of intercepts is more

Table 32.5 Comparison of three BMI linear growth models with different definitions of time

Fixed effects	*Model 1 – Interview wave*				*Model 2 – Age*				*Model 3 – Age12*			
	Coef.	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*
Intercept (γ_{00})	21.72	0.047	463.9	0.000	14.12	0.108	131.1	0.000	20.63	0.050	413.7	0.000
Linear (γ_{10})	0.58	0.007	83.0	0.000	0.54	0.006	84.6	0.000	0.54	0.006	84.6	0.000
Model fit												
Deviance	261356.6				260895.1				260895.1			
Parameters	6				6				6			
AIC	261368.6				260907.1				260907.1			
BIC	261422.0				260960.6				260960.6			

meaningful. Remember that an intercept is the predicted value of a dependent variable when the predictors are all 0. Consider Models 2 and 3 in Table 32.5. The intercept for BMI in Model 2 is 14.12. We interpret this as the predicted value of BMI when a person is 0 years old. This interpretation is not useful—BMIs for infants are not defined or interpretable! If we center age by subtracting 12 from each score, the only change in the model is that the intercept is now 20.63. This is the predicted value for a person who is 12 years old (Age $-12 = 0$). This is much more meaningful, because it represents the value for the youngest persons who were actually included in the NLSY97 study. If we had centered age by subtracting the grand mean of age across all of the participants and time points, we would have a different intercept. This grand-mean centered intercept would be interpreted as the expected BMI score for a person who was the average age of all persons in the study. For growth models it is fairly typical to center the time variable by subtracting the time at the first observation. This allows an interpretation of the intercept as the "starting point" of the growth curve.

The second reason that centering is typically done in growth models has to do with problems of multicollinearity. If polynomial transformations of time are used to build nonlinear growth models, the various time predictors (time, time-squared, etc.) are highly intercorrelated, and may lead to convergence problems, especially with smaller datasets. By centering all of the time variables the intercorrelations are reduced, and convergence problems will be less likely.

3.6 Identifying predictors of change

The above presentation has focused on developing growth curve models whose primary goal is to describe the form or shape of change. Typically, however, researchers are also interested in developing models and testing hypotheses that include predictors of change. In addition to the level 1 Time variable, growth curve models can include other types of level 1 (intraindividual) and level 2 (interindividual) covariates or predictors.

3.7 Predictors of interindividual change

In growth curve models, covariates that are constant over time, such as gender or experimental condition, are known as interindividual predictors. These predictors can tell us how change varies across different types of individuals. For our example, we will examine the effects of gender and ethnicity on the change in Substance

Use Days. To start we examine whether gender (Male) and ethnicity (NonWhite) affect the intercept of Substance Use Days (SUD). Both of these predictors are binary, with 1 indicating male or nonwhite, respectively. (The original NLSY97 ethnicity variable was categorical with several ethnicity options. This was recoded to 0=white, 1=nonwhite for this example.) The following equation shows the growth model to be fitted.

$$SUD_{ti} = \beta_{0i} + \beta_{1i}(Age12)_{ti} + \varepsilon_{ti}$$

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(Male)_i + \gamma_{02}(NonWhite)_i + u_{0i}$$

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

This growth model makes it clear that Male and NonWhite are entered as level 2 predictors for the intercept (β_{01}) of Substance Use Days. Note that this means for this first model that we assume a single linear slope for SUD on Age.

The results of fitting this first predictor model are shown in the left-hand side of Table 32.6. Both the gender and ethnicity predictors are highly significant. The intercept (1.39) is now interpreted as the predicted number of Substance Use Days for a white female age 12. The gender effect (1.17) tells us that 12-year-old males use substances about one day a month more often, and the ethnicity effect (−4.15) tells us that nonwhites are much less likely to use substances when they are young.

However, it may be that the relationship between age and substance use may not be the same across the different gender and ethnic groups. To test this, we can fit a more complex model that allows the linear slope of

Table 32.6 Effects of gender and ethnicity on change in Substance Use Days

Fixed effects	*Model 1 – Intercept effects*				*Model 2 – Slope and intercept effects*			
	Coef.	*SE*	*t-ratio*	*p*	*Coef.*	*SE*	*t-ratio*	*p*
Intercept (γ_{00})	1.392	0.242	5.7	0.000	1.708	0.290	5.9	0.000
Male (γ_{01})	1.170	0.228	5.1	0.000	−1.690	0.321	−5.3	0.000
NonWhite (γ_{02})	−4.150	0.226	−18.4	0.000	−1.722	0.319	−5.4	0.000
Age12 slope (γ_{10})	1.661	0.030	55.6	0.000	1.582	0.050	31.8	0.000
Male (γ_{11})					0.738	0.058	12.6	0.000
NonWhite (γ_{12})					−0.624	0.059	−10.6	0.000
Random effects	*Variance component*		χ^2	*p*	*Variance component*		χ^2	*P*
Intercept (u_{0i})	84.62		15280	0.000	79.53		15000	0.000
Linear Slope (u_{1i})	4.27		21241	0.000	4.03		20699	0.000
Level 1 (ε_{ti})	100.07				100.24			
Model fit								
Deviance	430916.6				430648.0			
Parameters	8				10			

SUD on Age12 to vary by gender and ethnicity. This corresponds to the following growth model:

$$SUD_{ti} = \beta_{0i} + \beta_{1i}(Age12)_{ti} + \varepsilon_{ti}$$
$$\beta_{0i} = \gamma_{00} + \gamma_{01}(Male)_i + \gamma_{02}(NonWhite)_i + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}(Male)_i + \gamma_{12}(NonWhite)_i + u_{1i}$$

Here we can see that gender and ethnicity are allowed to affect not only the intercept of SUD, but also the slope of SUD on Age. This model can be re-expressed in the mixed effects format as:

$$SUD_{ti} = \gamma_{00} + \gamma_{01}(Male)_i + \gamma_{02}(NW)_i + \gamma_{10}(Age)_{ti} + \gamma_{11}(Male)_i(Age)_{ti} + \gamma_{12}(NW)_i(Age)_{ti} + u_0 + u_1 + \varepsilon_{ti}$$

Although somewhat complicated, the mixed effects version highlights the fact that by including level 2 predictors of the slope we are actually entering cross-level interactions into the model. For example, γ_{11} will assess the extent to which the slope of SUD on age varies between girls and boys. This is, in effect, an interaction between gender (level 2) and time (level 1). Cross-level interactions are often the effects of most interest to researchers. For example, in longitudinal clinical trials, the test of the effectiveness of an intervention is typically modeled as a cross-level interaction of experimental condition (e.g., experimental vs control groups) by time.

The results of this model can be seen in the right side of Table 32.6. The parameter estimates show that gender and ethnicity are strong predictors of both the intercept and slope of time. To interpret these effects, it again helps to plot the prediction equations (Figure 32.5). Here we see the same basic finding that substance use increases over time. However, this model also shows us that we expect this increase to be the greatest for white males, and the lowest for nonwhite females.

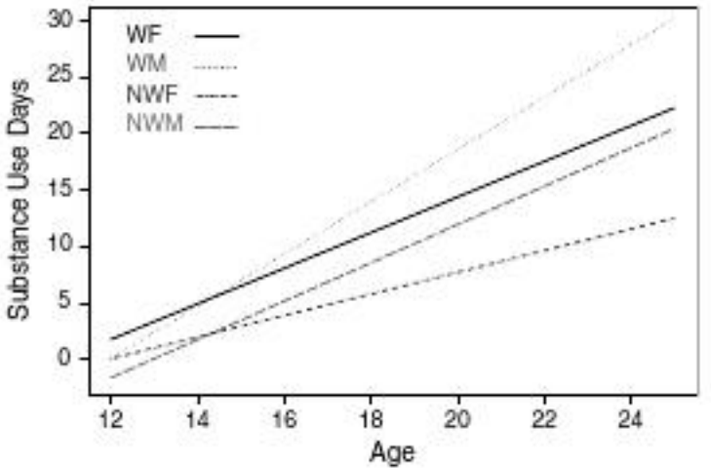

Figure 32.5 Predicted effects of gender and ethnicity on linear change of total Substance Use Days

3.8 Predictors of intraindividual change

There are many times that the predictors of interest in growth curve modeling will change over time themselves. These time-varying covariates can be used to model intraindividual change. Consider, for example, the effects of school transitions on daily substance use among youth. Social scientists have viewed school transitions as times of higher stress for students, as well as opportunities to form new social networks, have more freedom from previous family and peer expectations, and otherwise provide a changed social environment for substance use. It would be reasonable to assume that substance use may look different after a school transition. How could this be modeled?

First, consider Table 32.7, which shows an example data file that could be used in growth curve modeling. Male is a variable denoting gender that is a constant covariate—it does not change over time during the study. College transition, on the other hand, is an indicator variable that is 1 when an interviewee is attending a new college during the time of the interview. This is a time-varying covariate—it can take on different values (although only two values for a binary variable) at different time points.

Table 32.7 Example data file with constant and time-varying covariates

ID	*Interview wave*	*Age*	*Substance Use Days*	*Male*	*College transition*
001	1	13	0	1	0
001	2	15	3	1	0
001	3	16	8	1	0
001	4	16	4	1	0
001	5	18	10	1	1
001	6	19	12	1	0
001	7	20	10	1	0
002	1	15	2	0	0
002	2	16	5	0	0
002	3	16	8	0	0
002	4	18	15	0	1
002	5	19	14	0	0
002	6	20	12	0	1
002	7	21	12	0	0

Using the NLSY97 data, we can examine the effects of college transition on substance use over time using the following model:

$$SUD_{ti} = \beta_{0i} + \beta_{1i}(College)_{ti} + \beta_{2i}(Age12)_{ti} + \beta_{3i}(College)(Age12)_{ti} + \varepsilon_{ti}$$

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

$$\beta_{1i} = \gamma_{10}$$

$$\beta_{2i} = \gamma_{20} + u_{3i}$$

$$\beta_{3i} = \gamma_{30}$$

The college transition variable is entered into the level 1 part of the growth model, because it can take on different values at different time points (as suggested by the *t* subscript). College appears twice in the level 1 part of the model. The college main effect (β_{1i}) will assess the effects of college transition on the intercept of Substance Use Days. That is, it will allow us to see how much substance use shifts up or down during a year when there is a college transition. The college by age interaction term (β_{3i}), on the other hand, allows us to see if there is a change in the slope of substance use over time after a college transition. This model has no level 2 predictors, and note that only the intercept and Age12 are modeled as random effects. We assume for this example that the effects of college transitions are the same for all individuals.

The results of fitting this model are presented in Table 32.8. Similar to our previous models, we see that youth start out at age 12 using substances approximately 0 days, and that this increases by about 1.6 days of use per year. A transition to college is associated with an upward shift of 4.2 substance use days. However, after a college transition, the upward trend over time has been reduced by .47 days per year. This can be seen more clearly in the prediction graph in Figure 32.6. In this graph we examine the predicted growth curve of substance use days for a person who enters college at age 18. The vertical dashed line represents

Table 32.8 Effects of college transitions on change in Substance Use Days

Fixed effects	*Model 1 – Intercept effects*			
	Coef.	*SE*	*t-ratio*	*p*
Intercept (γ_{00})	−0.007	0.161	−0.0	0.967
College (γ_{10})	4.160	0.923	4.5	0.000
Age12 (γ_{20})	1.649	0.030	54.8	0.000
College X Age12(γ_{30})	−0.471	0.126	−3.7	0.000
Random effects	*Variance component*		χ^2	*p*
Intercept (u_{0i})	80.64		15083	0.000
Linear Slope (u_{1i})	4.26		21222	0.000
Level 1 (ε_{ti})	100.12			
Model fit				
Deviance	431225.1			
Parameters	8			

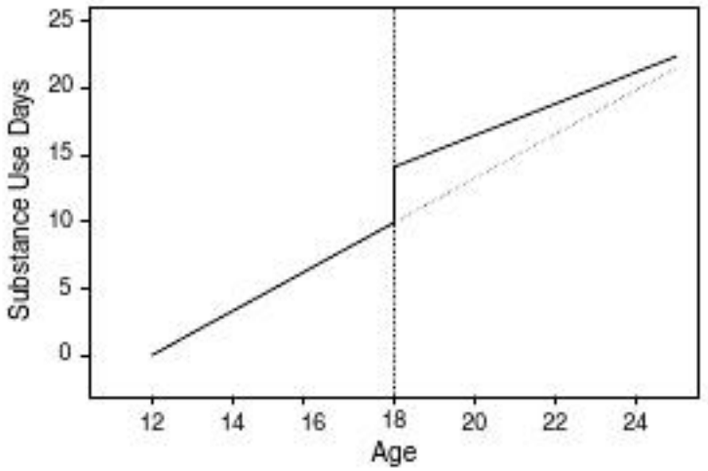

Figure 32.6 Predicted effects of college transition (at age 18) on linear change of Substance Use Days

the transition to college at age 18. At that point we see the sudden shift of substance use days upward to about 14 days. This may reflect the greater access to alcohol that many college students experience. After the transition to college, substance use still increases, but at a slower rate than for youth who do not make a transition to college (represented by the dashed line that continues upward to the right).

4 Conclusion

As we have seen, multilevel growth curve modeling is a flexible tool for analyzing longitudinal data. By viewing longitudinal data as observations nested within individual cases, we can use the power of multilevel modeling to answer questions about patterns and predictors of change. A number of advanced or more technical topics have been passed over or only mentioned briefly in this chapter. In particular, this chapter has focused on the use of growth curve modeling for quantitative dependent variables. Growth curve models can be built for other types of dependent variables, including binary, count, and ordinal variables. For more detailed treatment of these generalized multilevel models, see the relevant sections in Hox (2002) and Snijders and Bosker (1999). Also, the next chapter in this volume deals with multilevel change models for categorical dependent variables.

Software

Users have a large number of good choices for software for fitting growth curve models. Any statistics package that includes mixed effects modeling or multilevel modeling can be used to develop growth curve models of the type discussed in this chapter. Table 32.9 lists the major software packages that are widely known and are powerful enough to develop a wide variety of growth curve models. Users can choose to use specialized software that focuses primarily on multilevel modeling (i.e., HLM or MLwiN), or general-purpose statistical software that includes mixed effects modeling procedures (i.e., R/S-Plus, SAS, SPSS, or Stata). Users new to growth curve models may want to learn these procedures using the specialized software. The interface and documentation of these packages make for a shallower learning curve for growth curve modeling. More experienced users may wish to use the general-purpose software. In particular, the data management and graphical exploration features of R, SPSS, SAS, and Stata cannot be matched by HLM or MLwiN.

The Centre for Multilevel Modelling (sic) maintains a comprehensive list of reviews of software packages for multilevel modeling at: http://www.mlwin.com/softrev/index.html. All of the packages listed in Table 32.9 are reviewed at this site, but some of the reviews are out of date. An extremely useful site for learning about multilevel software is UCLA's statistical computing portal at: http://www.ats.ucla.edu/stat/. For example, all of the data and examples from Singer and Willett's textbook on longitudinal data analysis are presented for each of the six software packages listed in Table 32.9. See http://www.ats.ucla.edu/stat/examples/alda.htm.

Table 32.9 Information about growth curve modeling software

	Specialized multilevel modeling software			
	Version	*Interface*	*Information*	*Core references*
HLM	6.02	Graphical	http://www.ssicentral.com	Raudenbush, et al. (2000)
MLwiN	2.01	Graphical	http://www.mlwin.com	Rasbash, et al. (2000)
	General statistics software			
	Version	*Interface*	*Information*	*Core references*
R/S-Plus – nlme or lme4	R: 2.3.0; S-Plus: 7	Syntax	http://www.r-project.org/ http://www.insightful.com	Pinheiro and Bates (2000)
SAS – Proc MIXED	9.1.3	Syntax	http://www.sas.com	Singer (1998)
SPSS – MIXED	14	Either	http://www.spss.com	SPSS Advanced Models documentation
Stata – gllamm and xtmixed	9	Syntax	http://www.stata.com http://www.gllamm.org	Rabe-Hesketh and Skrondal (2005)

Glossary

AIC Akaike Information Criteria—a parsimony corrected measure of model fit.

BIC Bayesian Information Criteria—a parsimony corrected measure of model fit.

Centering A reparameterization of a predictor by subtracting a grand mean, group mean, or constant.

Deviance This is –2 times the log-likelihood of an estimated model.

Fixed effect This corresponds to the constant effects across persons of a predictor variable in a growth curve model.

Growth curve model A mixed effects model applied to longitudinal data.

Maximum likelihood estimation The most common type of estimation technique used for growth curve models of quantitative dependent variables.

Mixed effects model A statistical model incorporating both fixed and random effects, useful for analyzing grouped and longitudinal data.

Random effect This corresponds to the variance components in a growth curve model. Parameters (slopes and intercepts) that are allowed to vary across persons are random effects.

References

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrica*, 52: 317–332.

Bryk, A. S. and Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101: 147–158.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F. and Alpert, A. (1999). *An Introduction to Latent Variable Growth Curve Modeling*. Mahway, NJ: Lawrence Erlbaum.

Hox, J. (2002). *Multilevel Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Laird, N. M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution.

Journal of the American Statistical Association, 73: 805–811.

Luke, D. A. (2004). *Multilevel Modeling.* Thousand Oaks, CA: Sage.

McArdle, J. J. and Nesselroade, J. R. (2002). Growth curve analyses in contemporary psychological research. In J. Schinka and W. Velicer (eds), *Comprehensive Handbook of Psychology,* Volume 2: *Research Methods in Psychology.* New York: Pergamon Press.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review,* 30: 66–85.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus.* New York: Springer.

Rabe-Hesketh, S. and Skrondal, A. (2005). *Multilevel and Longitudinal Modeling Using Stata.* College Station, TX: Stata Press.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. and Lewis, T. (2000). *A User's Guide to MLwiN.* London: University of London, Institute of Education.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F. and Congdon, R. (2000). *HLM 5: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood, IL: Scientific Software International.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics,* 6: 461–464.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics,* 24: 323–355.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press.

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advance Multilevel Modeling.* London: Sage.