A Primer in Longitudinal Data Analysis

Garrett M. Fitzmaurice, ScD; Caitlin Ravichandran, PhD

Longitudinal data, comprising repeated measurements of the same individuals over time, arise frequently in cardiology and the biomedical sciences in general. For example, Frison and Pocock[1] used repeated measurements of the liver enzyme creatine kinase in serum of cardiac patients to study changes in liver function over a 12-month study period. The main goal, indeed the *raison d'être*, of a longitudinal study is characterization of changes in the response of interest over time. Ordinarily, changes in the response are also related to selected covariates. For example, Frison and Pocock[1] compared changes in creatine kinase between patients randomized to active drug and placebo.

The past 25 years have witnessed remarkable developments in statistical methods for the analysis of longitudinal data. Despite these important advances, researchers in the biomedical sciences have been somewhat slow to adopt these methods and often rely on statistical techniques that fail to adequately account for longitudinal study designs. The goal of the present report is to provide an overview of some recently developed methods for longitudinal analyses that are more appropriate, with a focus on 2 methods for continuous responses: the analysis of response profiles and linear mixed-effects models. The analysis of response profiles is better suited to settings with a relatively small number of repeated measurements, obtained on a common set of occasions, whereas linear mixed-effects models are suitable in more general settings. Before describing these methods, we review some of the defining features of longitudinal studies and highlight the main aspects of longitudinal data that complicate their analysis.

Features of Longitudinal Studies

Covariance Structure

A common feature of repeated measurements on an individual is correlation; that is, knowledge of the value of the response on one occasion provides information about the likely value of the response on a future occasion. Another common feature of longitudinal data is heterogeneous variability; that is, the variance of the response changes over the duration of the study. These 2 features of longitudinal data violate the fundamental assumptions of independence and homogeneity of variance that are at the basis of many standard techniques (eg, *t* test, ANOVA, and multiple linear regression). To account for these features, statistical models for longitudinal data have 2 main components: a model for the covariance among repeated measures, coupled with a model for the mean response and its dependence on covariates (eg, treatment group in the context of clinical trials). By the term "covariance," we mean both the correlations among pairs of repeated measures on an individual and the variability of the responses on different occasions (conversely, correlation can be interpreted as the standardized covariance). Although the main scientific interest is normally focused on the model for the mean response, inferences about change in the response and its relation to covariates are sensitive to the chosen model for the covariance among the repeated measures. Failure to properly account for the covariance results in hypothesis tests and CIs that are invalid and may result in misleading inferences.

Balanced Versus Unbalanced Designs

Typically, longitudinal study designs call for a fixed number of repeated measurements on all study participants at a set of common time points. When all individuals have the same number of repeated measurements, obtained on a common set of occasions, the study is said to be "balanced" over time. Many of the early statistical methods developed for longitudinal analysis (eg, repeated-measures ANOVA) required that the data be balanced. However, in longitudinal studies in the health sciences, especially those with repeated measurements over a relatively long duration, some individuals almost always miss their scheduled visit or date of observation. In some studies, this may necessitate that observations be made before or after the scheduled visit; these are referred to as mistimed measurements. Consequently, the sequence of observation times is no longer common to all individuals. In that case, we call the data "unbalanced" over time.

Missing Data

When some observations are missing (eg, due to skipped assessments or study dropout), a ubiquitous problem in longitudinal studies, the data are necessarily unbalanced; however, to distinguish missing data in a longitudinal study from other kinds of unbalanced data, such data are often called "incomplete." This distinction emphasizes the fact that an intended measurement for

From the Laboratory for Psychiatric Biostatistics (G.M.F., C.R.), McLean Hospital, Belmont, Mass; Department of Biostatistics (G.M.F.), Harvard School of Public Health, Boston, Mass; and Department of Psychiatry (G.M.F., C.R.), Harvard Medical School, Boston, Mass.

The online-only Data Supplement is available with this article at http://circ.ahajournals.org/cgi/content/full/118/19/2005/DC1.

Correspondence to Garrett Fitzmaurice, Laboratory for Psychiatric Biostatistics, McLean Hospital, North Belknap 309, 115 Mill St, Belmont, MA 02478-9106. E-mail fitzmaur@hsph.harvard.edu

(*Circulation*. 2008;118:2005-2010.)

***Circulation* is available at http://circ.ahajournals.org** **DOI: 10.1161/CIRCULATIONAHA.107.714618**

an individual could not be obtained. Complete case analysis, a common and simple method for handling incomplete data, excludes individuals with 1 or more missing measurements from the analysis. Although sometimes required by less modern methods of longitudinal analysis, this approach can not only be highly inefficient when a large proportion of the subjects are excluded, it can also seriously bias estimates of longitudinal change when the individuals with complete data are not a random sample from the target population. Fortunately, most modern methods for longitudinal analysis accommodate incomplete data under less stringent assumptions. Although discussion of these assumptions and their associated ramifications goes beyond the scope of the present report, a thorough longitudinal analysis includes assessment of how reasonable these assumptions are for the data at hand and consideration of how their violation could impact the results of the analysis.

Overview of Longitudinal Analysis

A longitudinal analysis of within-individual change proceeds in 2 conceptually distinct stages. In the first stage, within-individual change is characterized in terms of some appropriate summary of the changes in the repeated measurements on each individual during the period of observation. An elementary measure of the observed within-individual change in the response is the change score or difference score. For example, differences between posttreatment and pretreatment measurements of the creatine kinase response provide measures of change in the liver enzyme. This very simple concept of within-individual change extends naturally from difference scores to more general response trajectories over time. For example, the difference score is proportional to the slope (or constant rate of change) of a linear response trajectory. Other kinds of response trajectories, for example, piecewise linear or curvilinear, can also parsimoniously summarize within-individual changes in the response over time.

In the second stage, these estimates of within-individual change are related to interindividual differences in selected covariates (eg, treatment group, smoking status, and gender). Although these 2 stages of the analysis are conceptually distinct, they can be combined within a single statistical model. Indeed, this is the basis of a very versatile class of models for longitudinal analyses of continuous responses known as linear mixed-effects models. Before discussing linear mixed-effects models, however, we describe a simpler and more traditional method for longitudinal analysis known as analysis of response profiles.

Analysis of Response Profiles

Our discussion of methods for longitudinal analysis begins with a balanced design, a common design for longitudinal clinical trials. In a longitudinal trial, subjects are randomly allocated to 1 of several treatments. Typically, in addition to measures taken after treatment commences, 1 or more measures occur before randomization. The measure taken before randomization is often called the baseline response.

In this setting, the analysis of response profiles provides a relatively straightforward method of analysis. Analysis of response profiles proceeds by comparing the sequence of mean responses over time among groups (ie, comparing their mean response profiles). If the pattern of change in the mean response is the same in the treatment groups, the response profiles should be parallel (ie, have the same overall shape). Formally, this translates into a statistical test of the null hypothesis of no interaction between treatment and time, where time is regarded as a categorical variable.

Historically, analysis of response profiles has been used extensively for the analysis of longitudinal data from designed experiments. Traditionally, it was implemented via either a univariate repeated-measures ANOVA or a multivariate ANOVA (MANOVA). In both approaches, the model for the mean is composed of group, time, and group-by-time interaction effects. However, univariate repeated-measures ANOVA makes the strong assumptions that the variances on each occasion are equal and that the correlations among pairs of repeated measures are equal, which are often unrealistic for longitudinal data. Although MANOVA places no constraints on the variances or correlations, its use may still be problematic, because many implementations require a complete case approach. Fortunately, modern methods for analysis of response profiles both allow flexible models for the covariance and accommodate incomplete data. In the following section, we describe an implementation using data on blood lead levels from a longitudinal clinical trial comparing an active treatment with placebo. A summary of advantages and disadvantages of the method follows the example.

Treatment of Lead-Exposed Children Trial

In the 1990s, a placebo-controlled, randomized trial of succimer, an oral chelating agent, was conducted in children exposed to lead.[2,3] The primary scientific question in this example was whether chelation treatment with succimer reduced blood lead levels over time relative to placebo or, equivalently, whether patterns of change from baseline differed between the 2 groups. The children were 12 to 33 months old at enrollment, with a mean blood lead level of 26 μg/dL at randomization. For illustrative purposes, we focus on a random subset of 100 children who participated in this trial and consider data on blood lead levels at baseline (or week 0), week 1, week 4, and week 6. The mean response profiles for the succimer and placebo groups are displayed in Figure 1, and the largest differences between the 2 treatment groups occurred during weeks 1 and 4 of the trial.

Results from the analysis of response profiles of the blood lead level data are presented in Table 1. Selected content of the data set and SAS code for implementing the analysis are presented in the Appendix (Data Supplement). The hypothesis tests for the main effects of the treatment and time factors are not of primary scientific interest. The primary scientific question of whether the 2 treatment groups differ in their patterns of change from baseline in mean blood lead levels translates directly into a test of the treatment group-by-time interaction. The test of the treatment group-by-time interaction yields a test statistic with 3 degrees of freedom (DF). Compared with the reference χ^2 distribution with 3 DF, this test provides strong evidence to reject the null hypothesis and conclude that the profiles are not parallel. Therefore, the patterns of change from baseline are not the same in the succimer and placebo groups.

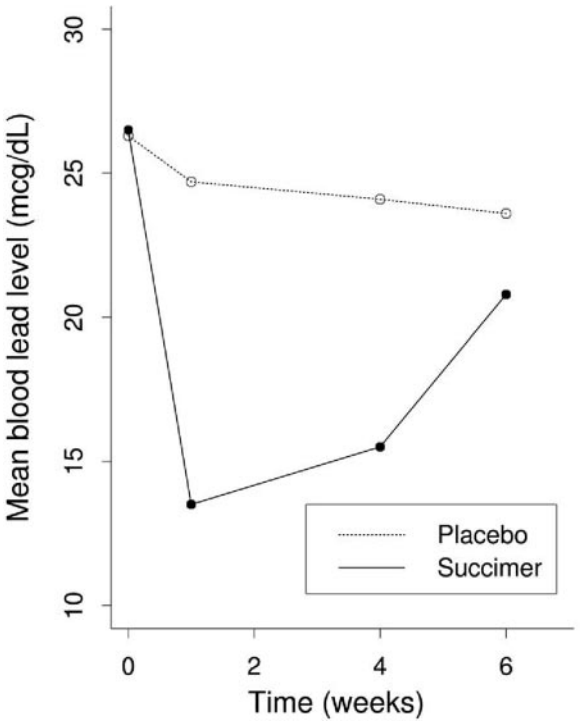

Figure 1. Plot of mean response profiles in the succimer and placebo groups.

Note that this omnibus test of interaction indicates that the 2 groups differ, but not how they differ. In cases in which a significant interaction is observed, comparison of group means at each measurement time can provide further information about the timing and persistence of group differences. Although space prohibits a display of the differences in means and SEs at each time point, additional hypothesis tests comparing changes from baseline at weeks 1, 4, and 6 indicated that children treated with succimer had a discernibly greater decrease in mean blood lead levels from baseline on all occasions than children treated with placebo. For example, compared with the placebo group, the succimer group had an additional 3.15 μg/dL (SE=1.26 μg/dL) decrease in mean blood lead levels from baseline to week 6. Of note, the largest differences between the 2 treatment groups occurred earlier in the trial. The rebound in blood lead levels after week 1 in the succimer group (Figure 1) was thought to be due to mobilization of the lead that is stored in the bones, which resulted in a new equilibrium in blood lead levels in the children treated with succimer.

Advantages and Disadvantages

One advantage of analysis of response profiles for researchers already familiar with ANOVA and basic regression techniques is that it can be conceptualized as an extension of ANOVA to the longitudinal setting. A second advantage is that because the method allows arbitrary patterns for the mean response over time and the covariance structure, the potential risks of bias due to model misspecification are minimal. However, analysis of response profiles has a number of potential drawbacks that make it either unappealing or unsuitable for analysis of data from many longitudinal studies. First, it is not well suited to handle mistimed measurements, a very common problem in many longitudinal studies. Second, the results of the analysis provide only a very broad or general statement about group differences in patterns of change over time. Ordinarily, a significant group-by-time interaction effect requires additional analyses to provide a more informative description of how the groups differ in their patterns of change. A related issue is that the omnibus test of group-by-time interaction may have relatively low power to detect group differences in situations in which change in the mean response over time can be summarized in a parsimonious way (eg, as a linear trend). Finally, the method requires estimation of a potentially large number of parameters in the models for the mean and covariance; in particular, the number of covariance parameters grows exponentially with the number of measurement occasions. As a result, the method will be more appealing in settings in which the number of subjects is relatively large compared with the number of measurement occasions.

Table 1. Results of an Analysis of Response Profiles of the Blood Lead Level Data at Baseline, Week1, Week 4, and Week 6 for the Children From the TLC Trial

Factor	DF	χ^2	*P*
Treatment	1	25.43	<0.0001
Time	3	184.48	<0.0001
Treatment-by-time	3	107.79	<0.0001

Linear Mixed-Effects Models

Like models for the analysis of response profiles, the versatile class of models known as linear mixed-effects models allows characterization and comparison of changes in the response of interest over time, complex models for the covariance, and accommodation of incomplete data; however, linear mixed-effects models can also handle unbalanced data, accommodate continuous covariates, and model the covariance in a parsimonious way. The underlying premise of linear mixed-effects models is that individuals in the population are assumed to have their own subject-specific mean response trajectories over time, ie, each individual has his or her own unique "curve" that describes longitudinal change in the response. This is accomplished by modeling the mean response as a combination of population characteristics that are assumed to be shared by all individuals and subject-specific effects that are unique to a particular individual. The former are called "fixed" effects, whereas the latter are called "random" effects. In this context, the term "mixed" denotes that the model contains both fixed and random effects.

These ideas are best understood with a simple but illuminating example. Consider a simple linear regression model for describing change in the mean response over time. In this common regression setting, we are interested in how the mean response changes over time for the population:

$$\underset{\text{Mean response at time j}}{E(Y_{ij}|\text{Time}_{ij})} = \underset{\text{Population intercept}}{\beta_0} + \underset{\text{Population slope}}{\beta_1^* \text{Time}_{ij}},$$

where Time_{ij} denotes the timing of the measure on the i^{th} individual at the j^{th} measurement occasion. In the above regression model, the population intercept, β_0, is the mean

response when Time=0 (perhaps denoting baseline), and the population slope, β_1, is the constant rate of change in the mean response for a single unit increase in time. Note that this model is a description of how the mean response changes in the population, where the averaging of the responses on any occasion is for all individuals. The mixed-effects model extends the above regression model by recognizing that individuals within a population are heterogeneous. Thus, at baseline (when Time=0), many individuals will have levels of response that are above or below the average, β_0. Similarly, many individuals will have a rate of change in the response that is more or less steep than the average rate of change, β_1. To accommodate both of these features, the mixed-effects model allows each individual to have his or her own subject-specific intercept and slope. This is accomplished by including not only the intercept and slope for the population (β_0 and β_1, respectively), as in the regression equation above, but also the intercept and slope (deviations) for each individual:

Y_{ij}	=	(β_0+b_{0i})	+	$(\beta_1+b_{1i})^*$ Time$_{ij}$	+	e_{ij}.
Response for individual i at time j		Intercept for individual i		Slope for individual i		Random error

In this model, β_0 is still the mean response, averaged across all individuals, at time zero; similarly, β_1 is the average rate of change across all individuals in the mean response over time. Here, β_0 and β_1 are called the fixed effects and describe the population-averaged response and how it changes over time. In addition, each random effect b_{0i} is the difference between the population-averaged intercept β_0 and the intercept for individual *i*, and each random effect b_{1i} is the difference between the population-averaged slope β_1 and the slope for individual *i*.

A graphic representation of the model is presented in Figure 2. The solid line represents the population-averaged response, which changes linearly over time. Individual A has a "higher" baseline response (when Time=0) than the population average (β_0) and therefore a positive b_{0i}. On the other hand, individual B has a "lower" baseline response and therefore a negative b_{0i}. In addition, individual A has a steeper rate of increase over time (β_1+b_{1i}) than the population average (β_1) and therefore a positive b_{1i} (see dotted line A). Individual B has a shallower rate of increase over time than the population average and therefore a negative b_{1i} (see dotted line B). Inclusion of the measurement errors, e_{ij}, allows the response on any occasion to vary randomly above and below the subject-specific trajectories. This random variation is reflected by the random scattering of points around the dotted subject-specific regression lines.

The model posits natural heterogeneity among individuals not only in terms of their initial level of response (when Time=0) but also in terms of changes in the response over time. The effects of covariates (eg, due to treatments, exposures, or background characteristics of the individuals) can be incorporated into the model for the mean, most often by including them as additional fixed effects. The introduction of random effects provides a flexible way to model the variability and correlation among repeated measures. When the model includes both random intercepts and slopes (or randomly varying coefficients for any functions of time), the variability of the response can change as a function of the times of measurement, and the magnitudes of the correlations between measurements from the same individual can depend on the time between them. In the illustration, we assume linear trends in the response over time, but this assumption could easily be relaxed (eg, by inclusion of a quadratic term for time in the model or via a nonlinear transformation of time, eg, log transformation). In the next section, we illustrate the use of linear mixed-effects models for longitudinal analysis using data from an observational study of pulmonary function decline and the effects of smoking.[4]

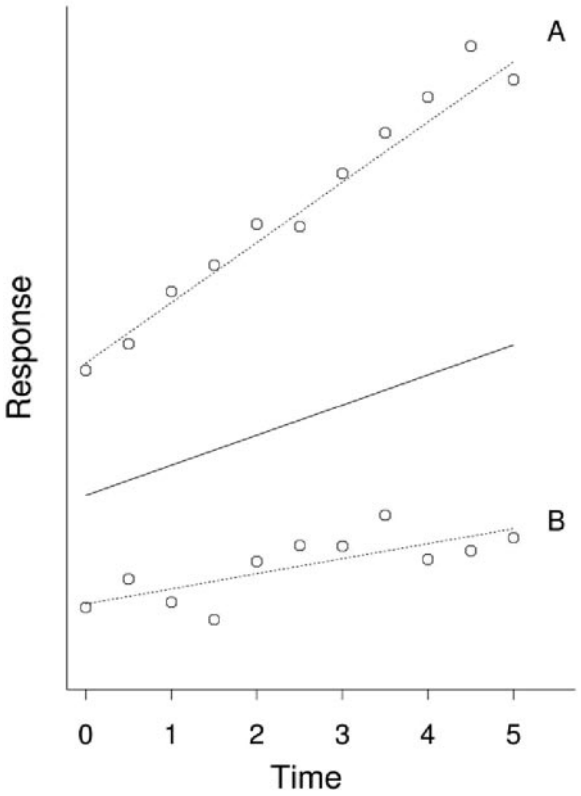

Figure 2. Graphic representation of mean response trajectory for 2 individuals (denoted A and B) and mean response trajectory averaged across all individuals.

The Vlagtwedde Study of Pulmonary Function

In an epidemiological study conducted in the rural area of Vlagtwedde, in the northeast part of the Netherlands, residents were followed up over a period of up to 21 years to obtain information on the prevalence of and risk factors for chronic obstructive lung diseases.[4,5] A measure of forced expiratory volume (FEV_1) was obtained every 3 years for the first 15 years of the study, as well as at year 19. In this study, FEV_1 was not recorded for every subject at each of the planned measurement occasions, so that the number of repeated measurements on each subject varied from 1 to 7. For the purpose of this illustration, we focus on a subset of the data on 133 current and former smokers 36 to 44 years of age at study entry whose smoking status did not change over the 19 years of follow-up. The goals of the analysis were to describe changes in lung function over the 19 years of

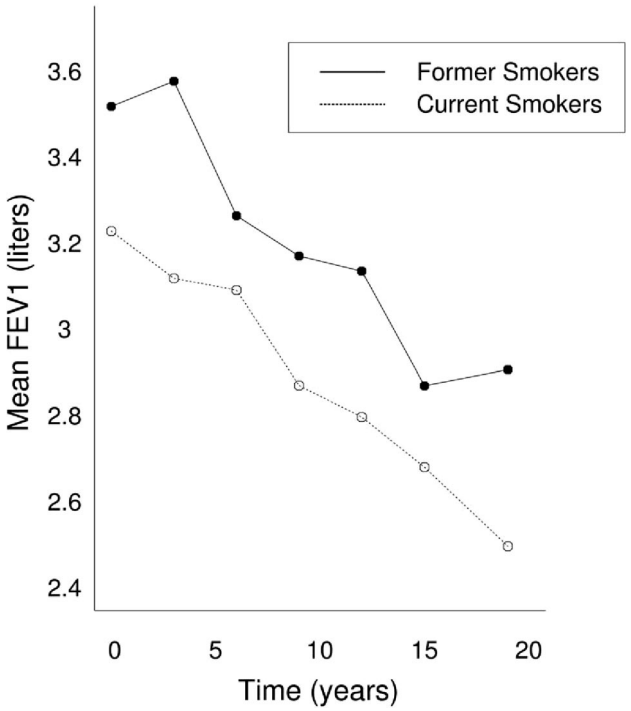

Figure 3. Plot of mean response profiles of FEV_1 at baseline (year 0), year 3, year 6, year 9, year 12, year 15, and year 19 in the current and former smoking exposure groups.

follow-up and to determine whether the pattern of change over time differed for current and former smokers.

The trends in mean FEV_1 over time for current and former smokers are displayed in Figure 3. Note that the pattern of decline was relatively constant over the duration of follow-up; consequently, we can approximate the pattern of change using linear trends. The comparison of trends in the 2 exposure groups can then be made by comparing the slopes of current smokers and former smokers. Specifically, we allow individuals to have randomly varying intercepts and slopes with means determined by smoking group. This corresponds to the following linear mixed-effects models, with randomly varying intercepts and slopes, for the 2 smoking exposure groups:

Former smokers:

$$FEV_{1ij} = (\beta_0+b_{0i}) + (\beta_1+b_{1i})^* Time_{ij} + e_{ij}$$

Current smokers:

$$FEV_{1ij} = (\beta_0+\beta_2+b_{0i}) + (\beta_1+\beta_3+b_{1i})^* Time_{ij} + e_{ij},$$

FEV_1 for individual i at time j	Intercept for individual i	Slope for individual i	Random error

where $Time_{ij}$ denotes the timing, in years, of the measure of FEV_1 on the i^{th} individual at the j^{th} measurement. Note that the expression for former smokers closely resembles the equation for a linear mixed-effects model presented earlier in the section, whereas the expression for current smokers contains 2 additional fixed effects, β_2 and β_3, which allow different population-averaged intercepts and slopes, respectively, for former smokers and current smokers. In this model, β_1 is the population mean change in FEV_1 per year for the former smoking group; the corresponding rate of change in FEV_1 per year for the current smoking group is $\beta_1+\beta_3$. Therefore, β_3 is the population difference in the rates of change in mean FEV_1 per year between the smoking groups. The question of main interest centers on the magnitude of β_3.

Table 2. Estimated Regression Coefficients and SEs Based on Mixed-Effects Model With Linear Trends for the FEV_1 Data From the Vlagtwedde Study

Effect	Estimate	SE	Z	*P*
Intercept	3.5051	0.1014	34.58	<0.0001
Time	−0.0337	0.0031	−10.96	<0.0001
Smoke	−0.2712	0.1162	−2.33	0.0196
Smoke-by-time	−0.0046	0.0035	−1.29	0.1972

To fit the model given above, we note that an equivalent way of writing the model is:

$$FEV_{1ij} = \beta_0+\beta_1^* Time_{ij}+\beta_2^* Smoke_i+\beta_3^* Smoke_i^* Time_{ij}+ b_{0i}+b_{1i}^* Time_{ij}+e_{ij},$$

where $Smoke_i$ is an indicator of the smoking exposure groups, with $Smoke_i=1$ for current smokers and $Smoke_i=0$ for former smokers. Selected content of the data set and SAS code for implementing the analysis are presented in the supplemental Appendix. Given the estimates of the regression coefficients in Table 2, it would appear that both groups had a significant decline in mean FEV_1 over the duration of the study, but there was no statistically discernible difference between the 2 smoking exposure groups in their rate of change, because the smoke-by-time interaction (ie, the comparison of the 2 slopes) was not significant ($Z=-1.29$, $P\approx0.2$). Thus, although current smokers had poorer lung function, as evidenced by their significantly lower mean FEV_1 at the start of the study ($Z=-2.33$, $P<0.05$), this disparity did not appear to widen over the 19 years of follow-up. Finally, we note that the fitted model makes a strong assumption that the pattern of change over time is linear in both exposure groups; the validity of this assumption can be tested and appears to be adequate for these data.

Conclusions

Longitudinal studies allow the investigation and comparison of changes in the response of interest over time. Although longitudinal data have characteristics that complicate analysis, such as correlations among repeated measurements, heterogeneous variability, and missing responses, modern methods can account for the often complex covariance structure and accommodate incomplete data. Routines in software packages widely used in medical research, such as PROC MIXED in SAS, xtmixed in Stata, and MIXED in SPSS, implement such methods.

Two approaches to modeling continuous longitudinal data are the analysis of response profiles and linear mixed-effects models. Analysis of response profiles is most appropriate when the data are balanced and the sample size is large relative to the number of measurement occasions. In contrast, linear mixed-effects models provide greater flexibility for

analysis of longitudinal data by accommodating unbalanced data and mixtures of discrete and continuous covariates and by modeling the covariance among repeated measures with a relatively small number of parameters. We have presented 2 applications of these methods to data sets from the medical literature. For further details about these methods and additional examples, we recommend consulting a text on applied longitudinal analysis, eg, Singer and Willett,[6] Fitzmaurice et al,[7] Brown and Prescott,[8] or Rabe-Heskreth and Skrondal.[9]

Other methods are available for longitudinal studies in which the response is not continuous, for example, studies with repeated binary measurements. Some of these methods can be conceptualized as extensions of linear mixed-effects models that accommodate different types of responses (eg, see chapters 11 to 13 in Fitzmaurice et al,[7] chapters 3 to 4 in Brown and Prescott,[8] or part III of Rabe-Hesketh and Skrondal[9]). Other methods for noncontinuous outcomes can be conceptualized as extensions of popular univariate methods such as logistic regression with an additional component that models associations among responses from the same individual.

In summary, as a result of the developments in statistical methods over the past 25 years, investigators now have available an array of new tools for longitudinal data analysis that can accommodate many common features of longitudinal studies, including inherently unbalanced designs, missing data and mistimed measurements, mixtures of discrete and continuous covariates, and outcomes that are counts or binary responses.

Disclosures

None.

References

1. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992; 11:1685–1704.
2. Treatment of Lead-Exposed Children (TLC) Trial Group. Safety and efficacy of succimer in toddlers with blood leads of 20–44 μg/dL. *Pediatr Res*. 2000;48:593–599.
3. Rogan WJ, Dietrich KN, Ware JH, Dockery DW, Salganik M, Radcliffe J, Jones RL, Ragan NB, Chisolm JJ, Rhoads GG. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. *N Engl J Med*. 2001;344:1421–1426.
4. van der Lende R, Kok TJ, Reig RP, Quanjer PH, Schouten JP, Orie NGM. Decreases in VC and FEV1 with time: indicators for effects of smoking and air pollution. *Bull Eur Physiopathol Respir*. 1981;17:775–792.
5. Rijcken B, Schouten JP, Weiss ST, Speizer FE, van der Lende R. The relationship of nonspecific bronchial responsiveness to respiratory symptoms in a random population sample. *Am Rev Respir Dis*. 1987;136: 62–68.
6. Singer JD, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press; 2003.
7. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley; 2004.
8. Brown H, Prescott R. *Applied Mixed Models in Medicine*. 2nd ed. New York, NY: Wiley; 2006.
9. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. College Station, Tex: Stata Press; 2008.

KEY WORDS: statistics as topic ■ data interpretation, statistical ■ linear models ■ data collection

A Primer in Longitudinal Data Analysis

Garrett M. Fitzmaurice and Caitlin Ravichandran

Circulation. 2008;118:2005-2010
doi: 10.1161/CIRCULATIONAHA.107.714618
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7322. Online ISSN: 1524-4539