

## Random Regression Models: A Comprehensive Approach to the Analysis of Longitudinal Psychiatric Data<sup>1</sup>

Robert D. Gibbons, Ph.D.,<sup>2</sup> Donald Hedeker, M.A.,<sup>3</sup> Christine Waternaux, Ph.D.,<sup>4</sup> and John M. Davis, M.D.<sup>2</sup>

### Introduction

Longitudinal studies occupy an important role in modern psychiatric research. In these studies the same individuals are repeatedly measured on a number of clinical and/or biological variables over a series of timepoints. As an example, a longitudinal design is often used to determine whether a particular therapeutic agent can produce changes in clinical status over the course of an illness. Another application for the longitudinal study is to assess potential indicators of a change in the subject's clinical status—for example, the assessment of whether drug plasma level measurements indicate clinical outcome.

Although the importance of the longitudinal design in psychiatric research is great, the methods of analyzing these data are not always commensurate with the time and energy spent collecting such data. For example, in many drug plasma level-clinical response studies, although both drug plasma levels and clinical status measures are usually collected on a series of occasions, most of the reported results do not take advantage of the time-varying nature of these data. Instead, researchers typically compute a clinical change score (based on baseline and final clinical assessments) and a steady-state drug plasma measure (averaged over several plasma level determinations), and then correlate the two measures. One reason for this simplified approach is that missing observations in the data prohibit the use of the traditional repeated measures anal-

ysis of variance (ANOVA) models. Also, the modeling of time-varying covariates, such as plasma level measurements, is not always possible with the traditional repeated measures approach. Thus, researchers often simplify the time-varying design of their data by computing an average value or change score.

Recently, random regression models have been developed to model continuous (Bock 1983; Laird & Ware 1982) or dichotomous (Gibbons & Bock 1987; Stiratelli et al. 1984) repeated measurements in which characteristics of the data preclude the use of the traditional ANOVA models. Specifically, random regression models allow for the presence of missing data, time-varying or invariant covariates, and subjects measured at different timepoints. Additionally, whereas the traditional repeated measures approach estimates average change in a population, the random regression approach can also estimate individual change for each subject. This is particularly useful in the psychiatric setting, where a proportion of subjects may respond to therapy in quite different ways from the average response.

In addition to the above references, other illustrations and developments of random regression models can be found in Gibbons and colleagues (1982), Jennrich and Schluchter (1986), Laird and colleagues (1987), and Ware (1985).

### Review of Traditional Methods

A comparison of several methods for the analysis of longitudinal data is provided in Table 1. These methods include (a) fixed-effects ANOVA, (b) repeated measures ANOVA (Winer 1971), (c) endpoint analysis, and (d) multivariate growth curve models (Bock 1975).

The fixed-effects ANOVA is the usual ANOVA model. When applied to longitudinal data, the model makes the completely unreasonable assumption that the within-subject measurements (over time) are uncorrelated.

The so-called repeated measures, or mixed-model, ANOVA assumes that the variances and covariances are constant over time. This assumption, termed "compound symmetry," implies that the correlation between responses made on two adjacent time points is identical to the correlation between responses made at any two points in time, including, for example, baseline and posttreatment. This assumption is, of course, unreasonable because the correlation between responses that are temporally proximal is almost al-

<sup>1</sup>This research was supported by a grant from the MacArthur Foundation.

<sup>2</sup>University of Illinois, Chicago, IL.

<sup>3</sup>University of Chicago, Chicago, IL.

<sup>4</sup>Harvard University, Cambridge, MA.

Reprint requests: Dr. R.D. Gibbons, Biometric Laboratory, Illinois State Psychiatric Institute, 1601 W. Taylor Street, Chicago, IL 60612.

**TABLE 1.** Comparison of Some Models for the Analysis of Longitudinal Psychiatric Data.

	Random Regression Model	Repeated Measures ANOVA	Fixed- Effects ANOVA	End- point Analysis	Multivariate Growth Curve
Allows missing data	yes	yes	yes	no	no
Subjects measured at different time points	yes	no	no	yes	no
General correlation structure over time	yes	no <sup>a</sup>	no <sup>b</sup>	no <sup>c</sup>	yes
Fixed covariates	yes	yes	yes	yes	yes
Time-varying covariates	yes	yes	yes	no	no
Discrete data generalizations	yes <sup>d</sup>	yes <sup>e</sup>	yes <sup>f</sup>	yes <sup>g</sup>	no
Includes all available data	yes	yes	yes	no	no
Provides estimates of individual rates of change	yes	no	no	no	no

<sup>a</sup>Assumes compound symmetry (equal variances and correlations over time).

<sup>b</sup>Assumes independence over time.

<sup>c</sup>Usually post- vs. pre-comparison only.

<sup>d</sup>Gibbons and Bock (1987).

<sup>e</sup>Koch and Landis type models (Koch et al. 1977).

<sup>f</sup>Log linear models.

<sup>g</sup>McNemar's test.

ways greater than the correlation between responses that are temporally distal.

Endpoint analysis refers to the procedure in which a *t*-test or an ANOVA is performed on change scores, and each patient's final measurement is used as the posttreatment score. This procedure is perhaps the most terrifying of all, in that it is capable of introducing extreme bias and misleading tests of hypotheses and estimates of precision. For example, it is often the case that for a relatively severe disorder, many patients who receive placebo drop out early because they have not benefited from their treatment. In contrast, if the active drug condition is at least moderately effective, most patients will complete the study. However, if the analysis treats all subjects as if they had all completed the study, regardless of the actual length of their participation, the two samples are no longer comparable.

In this case, the effect of treatment is confounded with the length of participation in the study. Furthermore, we would expect the final placebo rating, which would often occur early in the study, to reflect increased severity over those patients (either placebo or active drug) who completed the study. The result is a magnification of the active treatment vs. placebo difference and an inflated false positive rate.

Finally, the multivariate growth curve model provides a statistically attractive solution to the problem of analyzing longitudinal psychiatric data by assuming a very general correlational structure for the errors of measurement. Unfortunately, the model requires complete information on all subjects and observations that are similarly time structured (i.e., all subjects are measured on the same occasions). These requirements are rarely met in longitudinal psychiatric studies.

In the following, the general random regression model is described and illustrated using data from a large double-blind longitudinal study of the efficacy of neuroleptics in the treatment of schizophrenia. Results of each statistical method are compared and contrasted, and we illustrate how the previously described limitations of the traditional approaches are overcome using a random regression model.

### Model Description

Consider the following model for the measurement *y* of subject *i* on occasion *t*:

$$y_{it} = \alpha + \beta t + \alpha_i + \beta_i t + \varepsilon_{it} \quad (1)$$

$$i = 1, 2, \dots, n \text{ subjects,}$$

$$t = 0, 1, \dots, n - 1 \text{ timepoints}$$

where

$y_{it}$  is the measurement for subject *i* on occasion *t*;

$\alpha$  is the overall population intercept;

$\beta$  is the overall population slope;

$\alpha_i$  is the intercept for subject *i*;

$\beta_i$  is the slope for subject *i*; and

$\varepsilon_{it}$  is an independent residual distributed normally with mean 0 and variance  $\sigma^2$ .

We also assume that the distribution of individual intercepts and slopes in the population is bivariate normal  $N(\mu, \Sigma)$ , with

$$\mu = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}.$$

This model can be referred to as a personal trend or change model since it represents the measurements of  $y$  as a function of time, both at the individual ( $\alpha_i$  and  $\beta_i$ ) and population ( $\alpha$  and  $\beta$ ) levels. The intercept parameters indicate the starting point, and the slope parameters indicate the degree of change over timepoints. Additionally, this model can assess the residual variance  $\sigma^2$ , intercept variance  $\sigma_{\alpha}^2$ , slope variance  $\sigma_{\beta}^2$ , and covariance of the intercept and the slope  $\sigma_{\alpha\beta}$ .

We can also include terms in the model for covariates that do not change over time (time invariant) and for covariates that vary across the measured timepoints (time varying). This model can then be written as

$$y_{it} = \alpha + \beta t + \alpha_i + \beta_i t + \gamma x + \delta x_t + \varepsilon_{it}, \quad (2)$$

where the additional parameters

$\gamma$  is the coefficient for the time invariant covariate  $x$ , and

$\delta$  is the coefficient for the time-varying covariate  $x_t$ .

As an example of the personal trend model with both time-varying and time-invariant covariates, consider the following illustration: a Hamilton Psychiatric Rating Scale for Depression (HAM-D) and imipramine (IMI) plasma level are measured at four timepoints, and the subject's sex is known. Then, for a given subject  $i$ , the above model would be represented in matrix form by

$$\begin{bmatrix} \text{HAM-D}_{i1} \\ \text{HAM-D}_{i2} \\ \text{HAM-D}_{i3} \\ \text{HAM-D}_{i4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \text{sex}_i \text{ IMI}_{i1} \\ \text{sex}_i \text{ IMI}_{i2} \\ \text{sex}_i \text{ IMI}_{i3} \\ \text{sex}_i \text{ IMI}_{i4} \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}.$$

Notice that, although the above model has been illustrated for subjects with complete data on four timepoints, complete data are not necessary for the random regression model. Also, the above model assumes that the measurements are made at weeks 0 (baseline) through 3, but by reformulating the time matrix, subjects with data at different timepoints could also be included. Although the imipramine metabolite desipramine (DMI) is not included in the above illustration of the model, the effect of DMI, as well as certain clinical subtypes (e.g., endogenous or psychotic), could also be examined in the analysis.

## Example

Consider the following data collected in the National Institute of Mental Health schizophrenia collaborative study on treatment-related changes in overall severity on the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr & Klett 1966). Item 79, "Severity of illness," was scored as

- 1 = normal, not at all ill
- 2 = borderline mentally ill
- 3 = mildly ill
- 4 = moderately ill
- 5 = markedly ill
- 6 = severely ill
- 7 = among the most extremely ill

The experimental design and corresponding sample sizes are displayed in Table 2. Inspection of Table 2 reveals that the longitudinal portion of the study is highly unbalanced in that there are large differences in the number of measurements made in the 6 weeks of treatment. This, however, is quite typical of longitudinal studies in psychiatry, due to patient dropouts and complications of the illness that may make systematic measurement difficult.

Results for the random regression model are displayed in Table 3. Inspection of Table 3 reveals the following. First, the average severity at baseline for the placebo group was a rating of 5.24, which is between "severely" and "markedly" ill. The precision of this estimate is approximately one decimal point (i.e., SEM = .09), and it is clearly differentiable from zero ( $p \leq .001$ ). None of the three active treatment groups differed significantly from the placebo group in terms of their initial severity. The active treatment group intercepts were 5.08 for chlorpromazine, 5.02 for fluphenazine, and 5.14 for thioridazine.

Second, the rate of change for the placebo group was  $\beta_0 = -.142$ , which reflects an improvement of .142 units per week, or a total 6-week improvement of .852 units. All three active treatment groups exhibited significantly increased improvement rates relative to the placebo group. The rates of change were .357

TABLE 2. Experimental Design and Weekly Sample Sizes.

Treatment Group	Sample Size at Week						
	0	1	2	3	4	5	6
Placebo	110	108	5	89	2	2	72
Chlorpromazine	110	108	3	96	4	5	87
Fluphenazine	114	108	2	100	2	2	89
Thioridazine	106	107	4	93	3	0	90

**TABLE 3.** Parameter Estimates, Standard Errors, and Probability Values for the Random Regression Model.

Parameter	Estimate	SE	$p <$
Placebo intercept ( $\alpha_0$ )	5.243	.087	.001
Placebo vs. chlorpromazine ( $\alpha_0 - \alpha_1$ )	-.159	.123	.200
Placebo vs. fluphenazine ( $\alpha_0 - \alpha_2$ )	-.220	.122	.072
Placebo vs. thioridazine ( $\alpha_0 - \alpha_3$ )	-.100	.124	.418
Placebo slope $\beta_0$	-.142	.028	.001
Placebo vs. chlorpromazine ( $\beta_0 - \beta_1$ )	-.215	.039	.001
Placebo vs. fluphenazine ( $\beta_0 - \beta_2$ )	-.263	.039	.001
Placebo vs. thioridazine ( $\beta_0 - \beta_3$ )	-.249	.039	.001
$\sigma_\alpha$	.669		
$\sigma_\beta$	.202		
$r_{\alpha\beta}$	.126		

units per week for chlorpromazine, .405 for fluphenazine, and .391 for thioridazine. These improvement rates translate into overall 6-week improvement ratings of 2.14 units for chlorpromazine, 2.43 units for fluphenazine, and 2.35 units for thioridazine. Taking the intercepts and slopes together, we obtain estimated 6-week average ratings of  $5.24 - .142(6) = 4.39$  for placebo (moderately to markedly ill),  $5.08 - .357(6) = 2.94$  for chlorpromazine (mildly ill),  $5.02 - .405(6) = 2.59$  for fluphenazine (borderline to mildly ill), and  $5.14 - .391(6) = 2.79$  for thioridazine (borderline to mildly ill). Subsequent analysis revealed that none of the three active treatment groups differed significantly from each other.

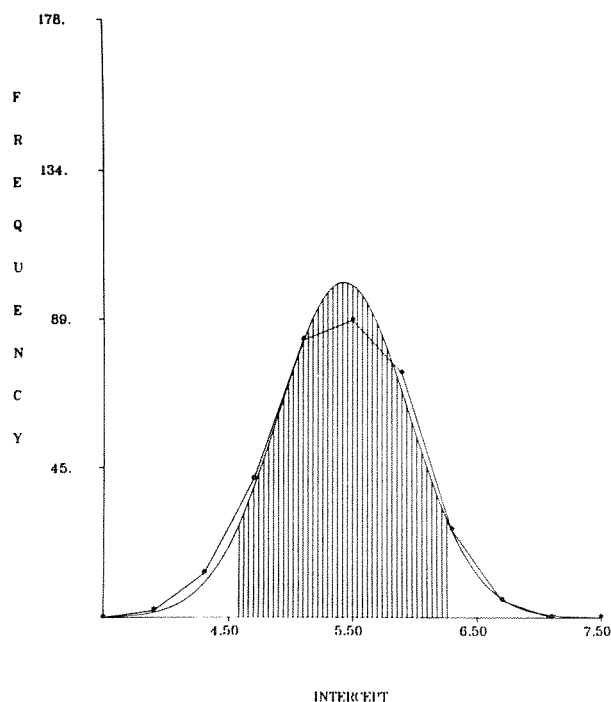
Third, the patient-specific effects revealed that at baseline, the standard deviation was on the order of  $\frac{3}{4}$  of a rating unit ( $\sigma_\alpha = .669$ ). Variability in rate of change was approximately .2 rating units per week ( $\sigma_\beta = .202$ ). Interestingly, the correlation between initial severity and rate of change was negligible ( $r_{\alpha\beta} = .126$ ), indicating that initial severity has little, if any, influence on the rate of a patient's recovery. Graphical representations of the probability distributions for these "personal effects" are displayed in Figures 1, 2, and 3. Figures 1 and 2 display the univariate marginal distributions for intercepts and slopes, respectively, for patients on active treatment. Figure 3 displays the joint bivariate normal distribution of intercepts and slopes (i.e., baseline severity and rates of change for patients on active treatment).

Table 4 presents some comparisons of the alternate models for the analysis of longitudinal data. Inspec-

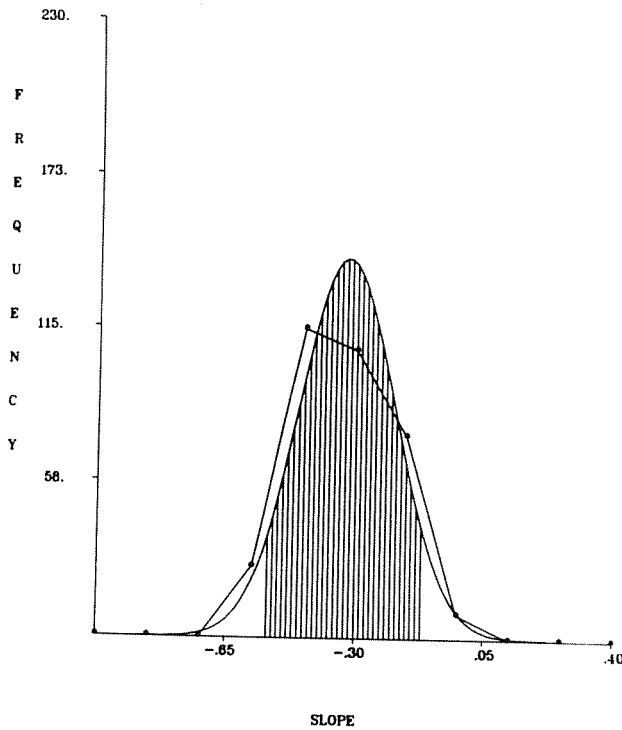
tion of Table 4 reveals that both the fixed-effects ANOVA and the multivariate procedure underestimated the significance of these differences (i.e., smaller  $z$ , larger  $p$ ), whereas the endpoint analysis overestimated the significance of the differences. The so-called repeated measures ANOVA could not even be calculated, even on a large mainframe computer, due to the large number of subjects. For this model, the design is group by subjects by time, which in this case consists of over 2,800 cells.

### Summary

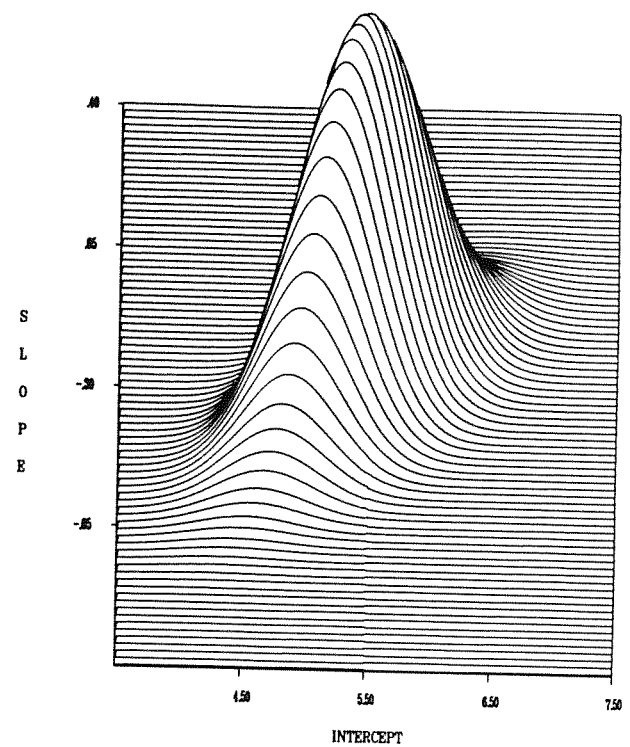
The random regression approach to the analysis of longitudinal psychiatric data provides numerous enhancements and flexibility over traditional approaches to this problem. Furthermore, the philosophical basis for this approach postulates that there is considerable heterogeneity in the response of individuals to treatment. It is the characterization of this heterogeneity and its relevance to the precision with which we can estimate the magnitude of treatment-related effects that make the random regression model a sensible



**FIGURE 1.** National Institute of Mental Health Collaborative Study—active treatment groups, Inpatient Multidimensional Psychiatric Scale item 79 (overall severity) random regression estimates: univariate marginal distribution for intercepts.



**FIGURE 2.** National Institute of Mental Health Collaborative Study—active treatment groups, Inpatient Multidimensional Psychiatric Scale item 79 (overall severity) random regression estimates: univariate marginal distribution for slopes.



**FIGURE 3.** National Institute of Mental Health Collaborative Study—active treatment groups Inpatient Multidimensional Psychiatric Scale item 79 (overall severity) random regression estimates: bivariate distribution of intercepts and slopes.

choice for the analysis of biobehavioral data. In contrast, the traditional approaches have often made the unfortunate and unrealistic assumption that individuals deviate randomly from some overall population-level process. Clearly, such an assumption is untenable for the measurement of human variation but not, perhaps, for the type of “split-plot” agricultural experiments for which they were designed.

**TABLE 4.** Between-Model Comparisons for the Placebo vs. Active Treatment Rates of Change (expressed as  $z = \text{estimate}/SE$ ).

Contrast	Random Regression Model ( $n = 446$ )	Fixed-Effects ANOVA ( $n = 1501$ )	End-point Analysis ( $n = 440$ )	Multivariate Growth Curve ( $n = 314$ )
Placebo vs. chlorpromazine	5.57	4.27	6.50	4.86
Placebo vs. fluphenazine	6.82	5.45	7.81	6.39
Placebo vs. thioridazine	6.43	4.80	7.43	5.49

NOTE: Critical value  $z = 1.96$ ,  $p \leq .05$ .

## References

- Bock, R.D. *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill, 1975.
- Bock, R.D. Within-subject experimentation in psychiatric research. In: Gibbons, R.D., and Dysken, M.W., eds. *Statistical and Methodological Advances in Psychiatric Research*. New York: Spectrum Press, 1983.
- Gibbons R.D., and Bock, R.D. Trend in correlated proportions. *Psychometrika* 52:113-124, 1987.
- Gibbons, R.D.; Clark, D.C.; and Davis, J.M. A statistical model for the classification of imipramine response in depressed inpatients. *Psychopharmacology* 78:185-189, 1982.
- Jennrich, R.I., and Schluchter, M.D. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42:805-820, 1986.
- Koch, G.G.; Landis, J.R.; Freeman, J.L.; Freeman, D.H.; and Lehnen, R.G. A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics* 33:133-158, 1977.
- Laird, N.M.; Lange, N.; and Stram, D. Maximum likelihood computations with repeated measures: Application of the EM algorithm. *J. Am. Stat. Assoc.* 82:97-105, 1987.

- Laird, N.M., and Ware, J.H. Random-effects models for longitudinal data. *Biometrics* 38:963-974, 1982.
- Lorr, M., and Klett, C.J. *Inpatient Multidimensional Psychiatric Scale: Manual*, revised. Palo Alto, CA: Consulting Psychologists Press, 1966.
- Stiratelli, R.; Laird, N.M.; and Ware, J.H. Random-effects models for serial observations with binary response. *Biometrics* 40:961-971, 1984.
- Ware, J.H. Linear models for the analysis of longitudinal studies. *Am. Statistician* 39:95-101, 1985.
- Winer, B.J. *Statistical Principles in Experimental Design*. 2d. ed. New York: McGraw-Hill, 1971.