

## Investigating Drug Plasma Levels and Clinical Response Using Random Regression Models

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

### Introduction

In the mid 1960s, researchers observed that individuals given the same oral dose of a drug exhibited variation in their steady-state plasma levels of that drug. Since this discovery, psychiatrists have been aware of the possibility of relating changes in drug plasma concentrations with changes in clinical response. While many investigations have been devoted to uncovering relationships between drug plasma levels and therapeutic effect, the statistical methods employed have not always allowed a rigorous examination of the data.

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 (which is 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 analysis of variance 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 where characteristics of the data preclude the use of the traditional analysis of variance models. Specifically, random regression models allow for the presence of missing data, time-varying or invariant covariates, and subjects measured at different time-points.

In a previous paper (Gibbons et al. 1988), we highlighted some of the differences between the random regression approach to longitudinal data and the more common approaches: endpoint analysis and the repeated measures analysis of variance models. A primary difference is that with the random regression model the focus of the analysis remains on the individual rather than on the sample group to which the subject belongs. For example, the analysis of variance models and the endpoint analysis compute statistics that indicate how the population or a subpopulation improves over time, whereas the random regression model also computes statistics for each individual that express how that particular subject is improving over time. Also, whereas in the more traditional approaches, the group means, standard deviations, and sample sizes are all that are necessary to perform the analysis, the random regression approach uses each subject's data throughout the estimation phase, iterating between the computation of the individual subject parameters and the overall population parameters. The individual parameters can then be analyzed to determine whether, for instance, the distribution of these parameters is unimodal or bimodal. A bimodal distribution might indicate the existence of treatment response subgroups rather than mere differences on a single continuum.

### Model Description

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

$$y_{it} = \alpha + \beta t + \alpha_i + \beta_i t + \epsilon_{it}$$

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

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

where

<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>Harvard University, Cambridge, MA.

Reprint requests: Donald Hedeker, Ph.D., Biometric Laboratory, Illinois State Psychiatric Institute, 1601 West Taylor Street, Chicago, IL 60612.

- 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
- $\epsilon_{it}$  is an independent residual distributed normally with mean 0 and variance

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

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

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 the covariance of the intercept and the slope  $\sigma_{\alpha\beta}$ . Note that the timepoints range from  $t = 0$  (i.e., the start of the study) to  $t = n_i - 1$ . The subscript  $i$  reveals that each subject may vary in terms of the number of occasions on which a response was recorded.

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, for a given subject  $i$ , as

$$y_{it} = \alpha + \beta t + \alpha_i + \beta_i t + \gamma x_i + \delta x_{it} + \epsilon_{it}$$

where the additional parameters

- $\gamma$  is the coefficient for the time invariant covariate  $x_i$ ; and
- $\delta$  is the coefficient for the time varying covariate  $x_{it}$ .

Interactions of the covariates can be included in the same way that interactions are included into a multiple regression model. One type of interaction that is particularly interesting is a time ( $t$ ) by covariate ( $x$  or  $x_i$ ) interaction. This type of interaction would indicate whether the covariate is related to the dependent measure differentially across timepoints.

For example, the effect of drug on clinical improvement could be increasing across the time period of the study.

The model that has been presented has thus far assumed that once we have conditioned on the individual (i.e., modeled the person-specific effects), the residuals are independent over timepoints. However, since the same subjects are measured repeatedly, some type of time-related dependency in the residuals must be expected. A simple type of dependency common for repeated measures data is a first-order autoregressive process. Regression models where the residuals follow an autoregressive process have been developed mainly in the econometric literature, for example, to model longitudinal earnings data (MaCurdy 1982). In this process, the residuals from successive timepoints are more highly correlated than the residuals from timepoints that are further apart. This pattern of correlation is sometimes referred to as serial correlation or autocorrelation. In terms of the model, the first order autoregressive process is written as, for subject  $i$  at timepoint  $t$ ,

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + e_{it}$$

where  $e_{it}$  is assumed to be normally distributed with mean 0 and variance  $\sigma^2$  and  $\rho$  is the autocorrelation coefficient. This equation simply states that the residuals at any timepoint are related to the residuals from the previous timepoint, and the magnitude of this relationship is represented by  $\rho$ . Notice that if  $\rho = 0$  there is no autocorrelation and the residuals are independent over timepoints. Using the *stationarity* assumption (Maddala 1977), that the residual variances and covariances depend only upon the intervals between the timepoints and not on the starting timepoint, the residual variance covariance matrix is of the form:

$$\Sigma_{\epsilon_i} = \frac{\sigma^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n_i-1} \\ \rho & 1 & \rho & \dots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n_i-2} & \rho^{n_i-3} & \dots & \dots & 1 \end{bmatrix}$$

The first subdiagonal of this matrix consists of the autocorrelation coefficient  $\rho$ , the second subdiagonal is  $\rho^2$ , the third is  $\rho^3$ , etc. Thus, the relationship between successive residuals is largest and the relationship decreases as the interval between the residuals increases. Also, this variance-covariance matrix is subscripted by the subject index  $i$  to indicate

that each subject can be measured on a different number of timepoints,  $n_i$ . This number then determines the size of the matrix ( $n_i \times n_i$ ).

The method of maximum likelihood, specifically Fisher's "method of scoring" solution (Silvey 1975), can be used to estimate the model and population parameters. This method also yields standard errors for each of the estimates that can be used to perform tests of significance, since the estimate divided by its standard error is essentially a z-score (Wald 1943). The computed z-score can be compared to a standard normal distribution frequency table to test the null hypothesis that the parameter equals 0, and when the z-score exceeds 1.96, the null hypothesis can be rejected at the usual  $p < .05$  critical level for a two-tailed test.

#### Example

Riesby and associates (1977) examined the relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients (37 endogenous and 29 nonendogenous). Following a baseline placebo period of 1 week, patients received 225 mg/day doses (75-mg tablets t.i.d) of IMI for the 4 weeks of the study. Blood samples, drawn twice weekly, 15 hours after the last drug intake of each patient, were assayed for IMI and DMI concentrations. Additionally, patients were rated using the Hamilton Rating Scale for Depression (HAM-D) during the baseline placebo period as well as at the end of each week of the study.

In their analysis of these data, Riesby and associates defined clinical response as a trichotomy (response, partial response, and nonresponse) based on the final HAM-D score of each patient. Steady-state plasma levels were computed as the mean value of the plasma concentrations at Weeks 2 through 4. The authors reported that all nonresponders (22) had steady-state plasma levels of IMI plus DMI below 240  $\mu\text{g/L}$ , while most of the responders in the endogenous group (10 of 12) had plasma levels above this limit.

While their results are informative, the time-varying nature of these data was not accommodated into their analysis. For instance, although the HAM-D was scored weekly during the study, only the final score was used in the data analysis. Their results indicate a relationship between average plasma levels and the final HAM-D score but do not indicate whether a

relationship exists throughout the course of the study. With random regression models, however, the time varying nature of the data can be included in the analysis, enabling the relationship between drug plasma levels and clinical response to be more fully explored.

The random regression model for a subject  $i$  can be represented in matrix form by

$$\begin{bmatrix} \text{HAM-D}_{i1} \\ \text{HAM-D}_{i2} \\ \text{HAM-D}_{i3} \\ \text{HAM-D}_{i4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} +$$

$$\begin{bmatrix} \text{sex}_i & \text{endog}_i & \text{IMI}_{i1} & \text{DMI}_{i1} \\ \text{sex}_i & \text{endog}_i & \text{IMI}_{i2} & \text{DMI}_{i2} \\ \text{sex}_i & \text{endog}_i & \text{IMI}_{i3} & \text{DMI}_{i3} \\ \text{sex}_i & \text{endog}_i & \text{IMI}_{i4} & \text{DMI}_{i4} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{bmatrix}$$

Although this model includes the plasma levels of IMI and DMI as well as the HAM-D scores for Weeks 1 through 4, complete data for all four timepoints are not required. In fact, while the random regression analysis used the data from all 66 subjects, the number of subjects with these 3 measures (HAM-D, IMI, and DMI) at each of the 4 timepoints fluctuated: 64 at Week 1, 65 at Week 2, 63 at Week 3, and 58 at Week 4.

The effects of sex and endogenous subtype are reflected by the regression coefficients  $\gamma_1$  and  $\gamma_2$ , while the coefficients  $\delta_1$  and  $\delta_2$  indicate the effects of IMI and DMI, respectively, on the HAM-D scores over the course of the study. Since the magnitude of the plasma level measurements varied greatly between individuals (from 4 to 312  $\mu\text{g/L}$  for IMI and from 0 to 740  $\mu\text{g/L}$  for DMI), a log transformation was used to ensure that the computed regression equation would not be unduly influenced by extreme, or outlying, values. Interactions of the plasma levels by time as well as the plasma levels by endogenous subtype were examined; however, these were not observed to be statistically significant and so were not included in the final model. Since the HAM-D ratings were made during the baseline placebo period in addition to the 4 weeks of the study, an analysis was performed on the actual HAM-D scores from Weeks 1 to 4 as well as the HAM-D change from baseline scores at Weeks 1 to 4.

The results for the random regression model, which are given in Table 1, indicate that the average HAM-D score at the end of Week 1 is a rating of 21.2.

TABLE 1 Random Regression Model Results.

Parameter	Estimate	SEM	p<
<i>HAM-D Total Score</i>			
Overall intercept $\alpha$	21.16	3.84	.0001
Overall slope $\beta$	-1.98	0.29	.0001
$\sigma_{\alpha}$	4.19	1.26	.001
$\sigma_{\beta}$	1.35	0.68	.05
$r_{\alpha\beta}$	0.14	0.30	NS
Sex $\gamma_1$	0.15	1.49	NS
Endogenous $\gamma_2$	2.33	1.29	.07
IMI $\delta_1$	0.44	0.83	NS
DMI $\delta_2$	-1.31	0.61	.05
Autocorrelation $\rho$	0.27	0.18	.07 <sup>a</sup>
<i>HAM-D Change From Baseline</i>			
Overall intercept $\alpha$	1.58	3.73	NS
Overall slope $\beta$	-1.93	0.29	.0001
$\sigma_{\alpha}$	3.85	1.23	.001
$\sigma_{\beta}$	1.40	0.66	.05
$r_{\alpha\beta}$	0.39	0.27	NS
Sex $\gamma_1$	-1.06	1.45	NS
Endogenous $\gamma_2$	1.28	1.26	NS
IMI $\delta_1$	0.51	0.81	NS
DMI $\delta_2$	-2.05	0.60	.001
Autocorrelation $\rho$	0.25	0.18	.08 <sup>a</sup>

NOTE: IMI = imipramine; DMI = desipramine; NS = nonsignificant.  
<sup>a</sup>One-tailed test of significance.

while the average improvement rate is roughly 2 units per week. This rate of improvement is observed in both the analysis of the total score ( $\beta = -1.98$ ) and the change from baseline score ( $\beta = -1.93$ ). The overall intercept ( $\alpha$ ) is nonsignificant in the change from baseline analysis, indicating that significant improvement was not evident between the baseline placebo period and the end of the first week of drug therapy (Week 1). Also, there does not seem to be a significant correlation between the intercept and the slope, since this parameter ( $r_{\alpha\beta}$ ) is nonsignificant in both analyses. Thus, the rate of improvement and the level of initial severity do not seem to be significantly related.

In terms of the covariates, the effects of sex and IMI are nonsignificant for both the HAM-D and the HAM-D change from baseline scores. The endogenous effect approaches significance for the HAM-D score, but is nonsignificant for the HAM-D change from baseline score. The effect of DMI is significant for both scores, although the effect is much stronger for the HAM-D change from baseline score. The correlations between the plasma levels and the HAM-D scores across the timepoints are given in Table 2. From these correlations, it is apparent that a consistent negative relationship exists between the

TABLE 2. Correlation Between HAM-D Scores and Plasma Levels (Natural Log Units).

Drug	Week 1	Week 2	Week 3	Week 4
<i>HAM-D Total Score</i>				
IMI	-0.034	-0.034	-0.003	-0.189
DMI	-0.178	-0.075	-0.250*	-0.293*
<i>HAM-D Change From Baseline</i>				
IMI	-0.025	-0.100	-0.034	-0.250
DMI	-0.350*	-0.274*	-0.348*	-0.401*

NOTE: IMI = imipramine; DMI = desipramine.  
 \* $p < .05$ .

HAM-D scores, especially the change from baseline scores, and the DMI plasma levels. Also, the relationship between IMI plasma levels and the HAM-D scores is not significant in the random regression model, and the correlations between these variables are small.

The residual autocorrelation is marginally significant, the actual magnitude ( $\rho = .25$ ) is moderate, and so the inclusion of this parameter in the model is warranted. The model residuals, thus, are not independent across timepoints and a moderate amount of autocorrelation is present. The significance test for this parameter is one-tailed, since we are only allowing the possibility of positive autocorrelation.

These data were also analyzed using both endpoint analysis and the repeated measures multivariate analysis of variance (MANOVA) model (Bock 1975). The univariate mixed analysis of variance model was not performed since the assumption of compound symmetry (equal variances and no correlation across timepoints) was rejected in these data. The endpoint analysis revealed a strong effect of Week 4 DMI values on Week 4 HAM-D change from baseline scores. This effect was comparable to that observed by the random regression model, although it only indicates that a relationship exists at Week 4, whereas the random regression model reveals a strong and consistent relationship for the entire 4 weeks of the study. Using the repeated measures MANOVA, which can assess whether a significant relationship exists over all timepoints, we did not observe a significant effect of DMI with either the HAM-D or the HAM-D change from baseline scores. Because this model requires complete data across timepoints for all variables, only 52 of the 66 subjects were used in the analysis. The significant DMI effect seen in the random regression analysis was not detected in this smaller sample by the repeated measures MANOVA.

### Summary

The reanalysis of the Riesby dataset using a random regression model indicates a significant effect of DMI plasma measurements on HAM-D scores across the four timepoints of the study. This effect is especially strong when the HAM-D change from baseline score is used in place of the actual HAM-D score at the four timepoints. A significant effect of IMI was not found for either the actual HAM-D score or the HAM-D change from baseline score. An endogenous effect was marginally significant when the actual HAM-D score was used; however, this effect was not observed when the HAM-D change score was used as the dependent measure. There also was evidence of a marginally significant effect due to autocorrelation of the residuals, indicating that the residuals at a given timepoint were related to the residuals from previous timepoints according to a first-order autoregressive process. In contrast, when the repeated measures MANOVA was used to analyze these data, a significant effect of DMI was not observed, since subjects without complete data at all timepoints had to be dropped from the analysis. In longitudinal psychiatric studies where missing data are the rule (rather than the exception), the random regression approach provides an attractive alternative to the traditional methods of analyzing longitudinal data.

### 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, 1983.
- Gibbons, R.D., and Bock, R.D. Trend in correlated proportions. *Psychometrika* 52:113-124, 1987.
- Gibbons, R.D.; Hedeker, D.; Waternaux, C.; and Davis, J.M. Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacol. Bull.* 24:438-443, 1988.
- Laird, N.M., and Ware, J.H. Random effects models for longitudinal data. *Biometrics* 38:963-974, 1982.
- MaCurdy, T.E. The use of time series processes to model the error structure in a longitudinal data analysis. *J. Econometrics* 18:83-114, 1982.
- Maddala, G.S. *Econometrics*. New York: McGraw-Hill, 1977.
- Riesby, N.; Gram, L.F.; Bech, P.; Nagy, A.; Petersen, G.O.; Ortmann, J.; Ibsen, I.; Dencker, S.J.; Jacobsen, O.; Krautwald, O.; Søndergaard, I.; and Christiansen, J. Imipramine: Clinical effects and pharmacokinetic variability. *Psychopharmacology (Berlin)* 54:263-272, 1977.
- Silvey, S.D. *Statistical Inference*. New York: Chapman and Hall, 1975.
- Stiratelli, R.; Laird, N.M.; and Ware, J.H. Random-effects models for serial observations with binary response. *Biometrics* 40:961-971, 1984.
- Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* 54:426-482, 1943.