Sankhyā : The Indian Journal of Statistics
Special Issue on Biostatistics
2000, Volume 62, Series B, Pt. 1, pp. 70–103

APPLICATIONS OF MIXED-EFFECTS MODELS IN BIOSTATISTICS*

By ROBERT D. GIBBONS
and
DONALD HEDEKER
University of Illinois at Chicago

SUMMARY. We present recent developments in mixed-effects models relevant to application in biostatistics. The major focus is on application of mixed-effects models to analysis of longitudinal data in general and longitudinal controlled clinical trials in detail. We present application of mixed-effects models to the case of unbalanced longitudinal data with complex residual error structures for continuous, binary and ordinal outcome measures for data with two and three levels of nesting (*e.g.*, a multi-center longitudinal clinical trial). We also examine other applications of mixed-effects models in the biological and behavioral sciences, such as analysis of clustered data, and simultaneous assessment of multiple biologic endpoints (*e.g.*, multivariate probit analysis). We describe the general statistical theory and then present relevant examples of these models to problems in the biological sciences.

1. Introduction

A common theme in the biological sciences is two-stage sampling *i.e.*, sampling of responses within experimental units (*e.g.*, patients) and sampling of experimental units within populations. For example, in prospective longitudinal studies patients are repeatedly sampled and assessed in terms of a variety of endpoints such as mental and physical level of functioning, or the response of one or more biologic systems to one or more forms of treatment. These patients are in turn sampled from a population, often stratified on the basis of treatment delivery such as a clinic, hospital, or community health system. Like all biological and behavioral characteristics, the outcome measures exhibit individual differences. We should be interested in not just the mean response pattern, but in the distribution of these response patterns (*e.g.*, time-trends) in the population of patients. Then we can speak of the number or proportion of patients who are functioning more or

---

ffig *AMS* (1991) *subject classification:* 92B15.
Key words and phrases: mixed models, longitudinal data, binary data, ordinal data, empirical Bayes, missing data.
* Supported by a Research Scientist Award from the National Institute of Mental Health grant K05-MH01254 to Dr. Gibbons and NIMH grants R01-MH44826 and R01-MH56146 to Drs. Hedeker and Gibbons.

less positively, at such and such a rate. We can describe the treatment-outcome relationship, not as a fixed law, but as a family of laws, the parameters of which describe the individual bio-behavioral tendencies of the subjects in the population (Bock, 1983). This view of biological and behavioral research leads inevitably to Bayesian methods of data analysis. The relevant distributions exist objectively and can be investigated empirically.

In biological research a very typical example of two-stage sampling is the longitudinal clinical trial in which patients are randomly assigned to different treatments and repeatedly evaluated over the course of the study. Despite recent advances in statistical methods for longitudinal research, the cost of medical research is not always commensurate with the quality of the analyses, often consisting of little more than an endpoint analysis in which only those subjects completing the study are considered in the analysis or the last available measurement for each subject is carried forward as if all subjects had, in fact, completed the study. In the first example of a completer only analysis, the available sample at the end of the study may have little similarity to the sample initially randomized. Things are somewhat better in the case of carrying the last observation forward, however, subjects treated in the analysis as if they have had identical exposure to the drug may have quite different exposures in reality or their experience on the drug may be complicated by other factors that led to their withdrawal from the study that are ignored in the analysis. Of course, in both cases there is a dramatic loss in statistical power due to the fact that the measurements made on the intermediate occasions are simply discarded. A review of the typical level of intra-individual variability of responses in these studies should raise serious question regarding reliance on any single measurement.

To illustrate the problem, consider the following example. Suppose a longitudinal randomized clinical trial is conducted to study the effects of a new pharmacologic agent for depression. Once a week, each patient is rated on a series of psychiatric symptoms (*e.g.*, depressed mood, trouble sleeping, suicidal thoughts, etc.) that may have some relation to the underlying disorder that the drug is intended to alleviate. At the end of the five-week study, the data comprise a file of number of symptoms rated positively for each patient in each treatment group. In addition to the usual completer and/or endpoint analysis a data analyst might compute means for each week and fit a linear or curvilinear trend line separately for each treatment group showing average number of symptoms per week. A more sophisticated analyst might fit the line using some variant of the Potthoff-Roy procedure, although this would require complete and similarly time structured data for all subjects (see Bock, 1979).

Despite the obvious question of whether the symptoms are equally related to the underlying disorder of interest (*e.g.*, depression) most objectionable is the representation of the mean trend in the population as a behavioral relationship acting within individual subjects. The analysis might purport that as any patient takes a given medication, he or she will decrease their number of positively rated symptoms at some fixed rate (*e.g.*, 3 symptoms per week). This is a gross oversimplification. The account is somewhat improved by reporting of mean trends for important subgroups *e.g.*, patients of high and low initial severity, males and females, and so on. Even

then within such groups some patients will respond more to a given treatment, some less, and others will not change at all. Like all behavioral characteristics, there are individual differences in response trends. Not only is the mean trend of interest, but so is the distribution of trends in the population of patients. Then we can speak of the number or proportion of patients who respond to a clinically acceptable degree and the rates at which their clinical status changes over time.

2. The General Linear Mixed-Effects Regression Model

Analysis of this type of two-stage data (under the assumptions that $\boldsymbol{\beta}$ has a distribution in the population of subjects, $\boldsymbol{\varepsilon}$ has a distribution in the population of responses within subjects and also in the population of subjects) belongs to the class of statistical problems called "mixed-model" (Elston & Grizzle, 1962; Longford, 1987), "regression with randomly dispersed parameters" (Rosenberg, 1973), "exchangeability between multiple regressions" (Lindley & Smith, 1972), "two-stage stochastic regression" (Fearn, 1975), "James-Stein estimation" (James & Stein, 1961), "variance component models" (Harville, 1977; Dempster, Rubin, & Tsutakawa, 1981), "random coefficient models" (DeLeeuw & Kreft, 1986), "hierarchical linear models" (Bryk & Raudenbush, 1987), "multilevel models" (Goldstein, 1986), and "random-effect regression models" (Laird and Ware, 1982). Along with these seminal articles, several book-length texts have been published further describing these methods (Bock, 1989a; Bryk & Raudenbush, 1992; Diggle, Liang, & Zeger, 1994; Goldstein, 1995; Jones, 1993; Longford, 1993; Lindsey, 1993). For the most part, these treatments are based on the assumption that the residuals, $\boldsymbol{\varepsilon}$, are similarly distributed as $N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ in all subjects and the personal trend parameters, $\boldsymbol{\beta}$, are $N(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$ independent of $\boldsymbol{\varepsilon}$.

To describe the model in a general way for data which are either clustered or longitudinal, the terminology of multilevel analysis can be used (Goldstein, 1995). For this, let i denote the level-2 units (clusters in the clustered data context, or subjects in the longitudinal data context), and let j denote the level-1 units (subjects in the clustered data context, or repeated observations in the longitudinal data context). Assume that there are $i = 1, \ldots N$ level-2 units and $j = 1, \ldots, n_i$ level-1 units nested within each level-2 unit. The mixed-effects regression model for the $n_i \times 1$ response vector $\boldsymbol{y}$ for level-2 unit i (subject or cluster) can be written as:

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\alpha} + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots N, \tag{1}$$

where $\boldsymbol{W}_i$ is a known $n_i \times p$ design matrix for the fixed effects, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of unknown fixed regression parameters, $n_i \times r$ design matrix for the random effects, $\boldsymbol{\beta}_i$ is the $r \times 1$ vector of unknown individual effects, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ error vector. The distribution of the random effects is typically assumed to be multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\beta$, and the errors are assumed to be independently distributed as multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\varepsilon = \sigma^2_\varepsilon\boldsymbol{\Omega}_i$. Although $\boldsymbol{\Omega}_i$ carries the subscript i, it

depends on i only through its dimension n_i, that is, the number of parameters in $\boldsymbol{\Omega}_i$ will not depend on i. In the case of independent residuals, $\boldsymbol{\Omega}_i = \boldsymbol{I}_i$, but for our purposes, we will define $\boldsymbol{\omega}$ to be the s x 1 vector of autocorrelation terms that $\boldsymbol{\Omega}_i$ depends on (Chi and Reinsel, 1989).

Different types of autocorrelated errors have been considered including first-order autoregressive process, AR(1), the first-order moving average process, MA(1), the first-order mixed autoregressive-moving average process, ARMA(1,1), and the general autocorrelation structure. A typical assumption in models with autocorrelated errors is that the variance of the errors is constant over timepoints and that the covariance of errors from differing timepoints depends only on the time interval between these timepoints and not on the starting timepoint. This assumption, referred to as the *stationarity* assumption, is assumed for the aforementioned forms. Another form of autocorrelated errors is described by Mansour, Nordheim, and Rutledge (1985), who examine autocorrelated errors which follow the first order autoregressive process, however, where the assumption of stationarity is relaxed.

As a result of the above assumptions, the observations $\boldsymbol{y}_i$ and random coefficients $\boldsymbol{\beta}$ have the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{\beta} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{W}_i\boldsymbol{\alpha} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\Sigma}_\beta\boldsymbol{X}_i' + \sigma_\varepsilon^2\boldsymbol{\Omega}_i & \boldsymbol{X}_i\boldsymbol{\Sigma}_\beta \\ \boldsymbol{\Sigma}_\beta\boldsymbol{X}_i' & \boldsymbol{\Sigma}_\beta \end{bmatrix} \right) . \tag{2}$$

The mean of the posterior distribution of $\boldsymbol{\beta}$, given $\boldsymbol{y}_i$, yields the empirical Bayes (EB) or EAP ("Expected A Posteriori") estimator of the level-2 parameters,

$$\tilde{\boldsymbol{\beta}}_i = \left[\boldsymbol{X}_i'(\sigma_\varepsilon^2\boldsymbol{\Omega}_i)^{-1}\boldsymbol{X}_i + {\boldsymbol{\Sigma}_\beta}^{-1}\right]^{-1} \boldsymbol{X}_i'(\sigma_\varepsilon^2\boldsymbol{\Omega}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\alpha}) , \tag{3}$$

with the corresponding posterior covariance matrix given by

$$\boldsymbol{\Sigma}_{\beta|y_i} = \left[\boldsymbol{X}_i'(\sigma_\varepsilon^2\boldsymbol{\Omega}_i)^{-1}\boldsymbol{X}_i + {\boldsymbol{\Sigma}_\beta}^{-1}\right]^{-1} . \tag{4}$$

Further details regarding estimation of $\boldsymbol{\Sigma}_\beta$, $\boldsymbol{\alpha}$, σ_ε^2 and $\boldsymbol{\omega}$ are provided in the Appendix.

2.0.1. *Illustration.* Gibbons *et al.*, (1993) reanalyzed the Hamilton Rating Scale of Depression (HRSD) data from the NIMH Treatment of Depression Collaborative Research Program (Elkin *et al.*, 1989). In the design of this study, a primary hypothesis involved the effectiveness of cognitive behavior therapy (CBT) and interpersonal psychotherapy (IPT) alone and as compared with each other in the treatment of outpatient depression. The major measure of depressive symptomatology was a modified version of the 17-item HRSD, completed by a "blind" clinical evaluator. As a standard reference treatment, an imipramine plus clinical management group (IMI-CM) was included in the design, and, as an additional control (particularly for the IMI-CM condition), a placebo plus clinical management group (PLA-CM). The study lasted 16 weeks and measurements were performed at weeks 0, 4, 12 and 16.

The original analysis of these data (Elkin *et al.*, 1989) compared the four treatment groups in terms of the last available HRSD measurement (*i.e.*, endpoint analysis) for all 239 subjects, 204 subjects having at least 3.5 weeks of treatment and

the 155 subjects that completed the trial. As expected, results of the analyses varied as a function of the sample analyzed. In the total sample of 239, the probability associated with the overall F-statistic for the comparison of the four groups approached significance ($p < .053$) with significant post-hoc comparisons of IMI vs PLA ($p < .017$) and IPT vs PLA ($p < .018$). No significant treatment related effects were seen for the other two samples (*i.e.*, $n = 204$ and $n = 155$). These equivocal results raised tremendous public controversy in that proponents of psychotherapy claimed that the results indicated equivalent benefit for pharmacotherapy and psychotherapy and used this as evidence to change insurance benefit plans.

Table 1 presents the mean baseline HRSD scores for those patients remaining in the study at each time point, as well the corresponding sample sizes. In general, patients that dropped out had slightly higher HRSD scores at baseline, but this effect was consistent for all four treatment groups.

Table 1. Mean Baseline HRSD Scores[1] for Sample Making it to Each Time-Point

	Week 0	Week 4	Week 8	Week 12	Week 16
All Centers					
CBT	19.6 (59)	19.7 (49)	19.4 (43)	19.1 (35)	19.2 (37)
IPT	19.6 (61)	19.2 (53)	18.8 (46)	18.9 (47)	18.9 (47)
IMI-CM	19.5 (57)	19.2 (48)	19.3 (44)	19.3 (39)	19.2 (37)
PLA-CM	19.5 (62)	19.0 (50)	19.0 (45)	19.1 (38)	19.1 (34)

[1] Re-screening scores
CBT = Cognitive Behavior Therapy
IPT = Interpersonal Psychotherapy
IMI-CM = Imipramine with Clinical Management
PLA-CM = Placebo with Clinical Management

Table 1 also shows that IPT had the lowest dropout rate (23%), PLA-CM had the highest (40%) whereas CBT (32%) and IMI-CM (33%) were intermediate and roughly the same.

Gibbons *et al.*, (1993) reanalyzed these data using a mixed-effects regression model applied to the total $n = 239$ sample. The general model posits that the individual response of each subject can be described by a line with intercept (baseline response) and slope (improvement rate) that is specific to the individual. Analysis of these data revealed that person-specific deviations in severity at baseline did not represent a significant component of variance, but variation in trend was significant; hence, a single random effect (*i.e.*, random trend model) was used. In this case, it was found that approximate linearity could be achieved by using a logarithmic transformation on time (*i.e.*, $\log_e$(weeks+1), see Figure 1 for observed means and Figure 2 for model fitted means).

First, it was found that in addition to a random trend effect, there was residual serial correlation best described by a first-order nonstationary autoregressive process with $\hat{\rho} = .35$, $p < .001$. Second, no significant differences were found between the two psychotherapies (IPT vs CBT), see Table 2. Third, imipramine produces a significantly faster rate of improvement relative to placebo ($p < .032$). The MMLE estimate of 1.2 HRSD units per log time (see Table 2), equals an average difference between IMI and PLA of 3.4 HRSD units at week 16. This effect can be seen in

Figure 1, where the observed raw response pattern for IMI-CM consistently draws away from that for PLA-CM. Figure 2 presents the model-fitted response patterns, in which the patterns (based on including all subjects at all times, and smoothing) are even clearer. Fourth, no significant differences were found between the two psychotherapies considered jointly (IPT + CBT = PSY) and the standard reference therapy (IMI-CM).

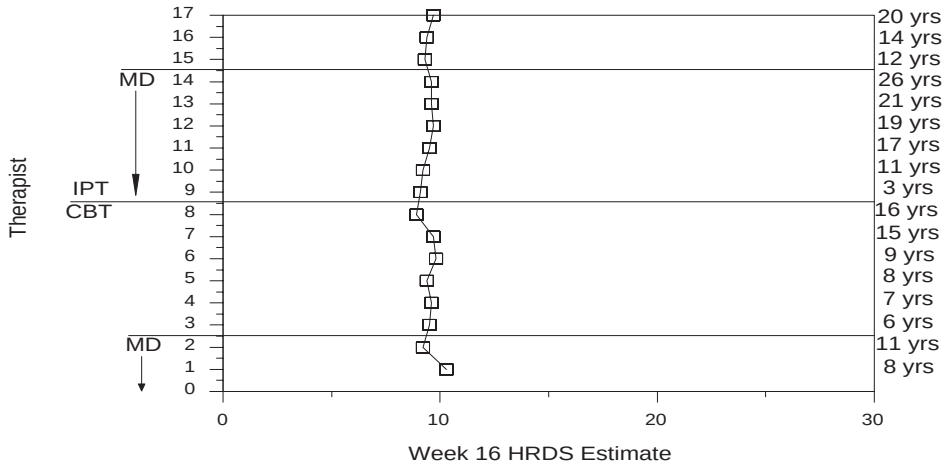

Figure 1. Observed HRSD group means for available sample at each month. The measured time points: weeks 4, 8, 12, and 16 correspond to x-axis values of 1.6, 2.2, 2.6, and 2.8 respectively.

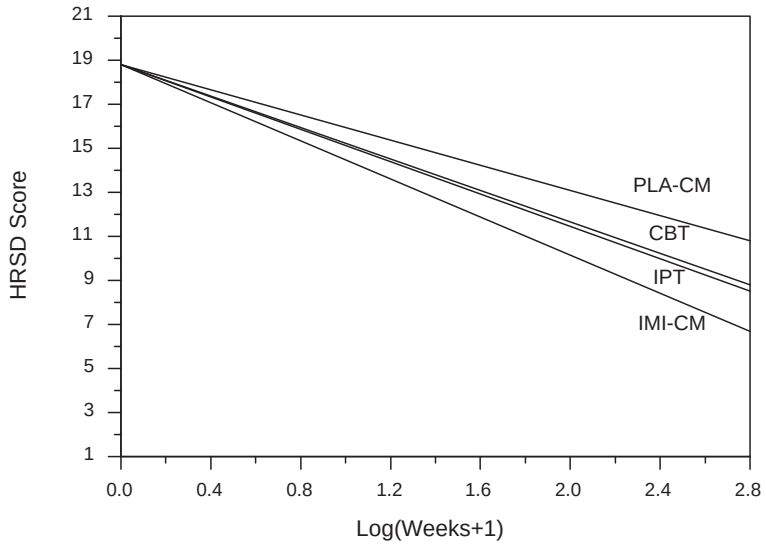

Figure 2. Estimated HRSD trend lines by treatment group using all available data. The measured time points: weeks 4, 8, 12, and 16 correspond to x-axis values of 1.6, 2.2, 2.6, and 2.8 respectively.

Table 2. Parameter Estimates, Standard Errors, Test Statistics
N = 239, Weeks 0 - 16, Log_e(Week + 1)

Effect	Estimate	SE	Z	P <
Overall Improvement Rate	-3.858	.196	-19.661	.001
Overall Baseline Response	19.349	.453	42.671	.001
Differences in Baselines:				
CBT vs IPT	0.049	.831	0.059	.953
PLA-CM vs IMI-CM	0.017	.835	0.020	.984
PSY vs IMI-CM	0.245	.732	0.335	.738
Differences in Improvement Rates:				
CBT vs IPT	0.388	.548	0.707	.480
PLA-CM vs IMI-CM	1.204	.561	2.145	.032
PSY vs IMI-CM	0.720	.484	1.486	.137
CBT vs IMI-CM	0.914	.563	1.621	.105
IPT vs IMI-CM	0.527	.549	0.960	.337
PLA-CM vs CBT	0.291	.560	0.519	.604
PLA-CM vs IPT	0.677	.546	1.240	.215

CBT = Cognitive Behavior Therapy
IPT = Interpersonal Psychotherapy
IMI-CM = Imipramine with Clinical Management
PLA-CM = Placebo with Clinical Management

No further significant differences among treatment groups were found, and this is apparent in Figure 2, the fitted response patterns. Inspection of Figure 2 reveals that PLA-CM patients exhibit the least response followed by the two psychotherapies. IMI-CM departs from the other trend lines early in treatment.

Note that in Figure 1, the raw response patterns show irregular behavior in the PLA-CM and IPT groups at 16 weeks, with what appears to be a sharp decrease in HRSD scores between 12 and 16 weeks. Gibbons *et al.*, (1993) redid the analysis, omitting the 16th week (see Table 3).

Table 3. Parameter Estimates, Standard Errors, Test Statistics
N = 239, Weeks 0 - 12, Log_e(Week + 1)

Effect	Estimate	SE	Z	P <
Overall Improvement Rate	-3.704	.215	-17.244	.001
Overall Baseline Response	19.348	.460	42.084	.001
Differences in Baselines:				
CBT vs IPT	0.099	.828	0.120	.905
PLA-CM vs IMI-CM	-0.460	.832	-0.055	.956
PSY vs IMI-CM	0.078	.729	0.107	.915
Differences in Improvement Rates				
CBT vs IPT	0.204	.601	0.339	.734
PLA-CM vs IMI-CM	1.489	.613	2.429	.015
PSY vs IMI-CM	1.318	.531	2.481	.013
CBT vs IMI-CM	1.419	.618	2.296	.022
IPT vs IMI-CM	1.215	.603	2.017	.044
PLA-CM vs CBT	0.070	.612	0.114	.909
PLA-CM vs IPT	0.273	.597	0.458	.647

CBT = Cognitive Behavior Therapy
IPT = Interpersonal Psychotherapy
IMI-CM = Imipramine with Clinical Management
PLA-CM = Placebo with Clinical Management

The differential effect of IMI-CM versus PLA-CM remained much the same ($p < .015$, a difference in improvement rate of 1.5 units per time, equal to an average difference of 4.2 HRSD units at week 12). Now, however, a significant difference between the two combined psychotherapy groups (PSY) and the Imipramine (IMI-CM) group was also found ($p < .013$, a difference in improvement rate of 1.3 units per time, equal to an average difference of 3.4 HRSD units at week 12), an effect clearly seen in Figure 1.

To illustrate the usefulness of person-specific effects in these models, the empirical Bayes estimates of several patients with missing measurements are displayed in Table 4. The highlighted values in Table 4 are the model estimates of the missing data and those values displayed in normal typeface are the observed values. The estimates of the missing measurements include the effects of the overall population intercept and trend, as well as treatment group differences and the random trend effect. The first four patients in Table 4 exhibited little or no response to treatment, the next three patients exhibited moderate response, and the final six patients exhibited a considerable response to treatment. The empirical Bayes estimates of the missing data appear to be reasonably consistent with the observed values, and are well differentiated between the three response groups displayed in Table 4. These estimates support the assumed ignorability of missing data in the standard mixed-effects regression model.

Table 4. Empirical Bayes Estimates of Missing Data

			week				
ID	group	slope	0	4	8	12	16
32	Pla	-1.15	17	23	17	20	**17.3**
57	Pla	-0.40	27	26	26	23	**19.8**
172	Cbt	-0.72	24	30	24	**19.9**	**17.5**
142	Pla	-2.70	14	16	**14.2**	**13.1**	**12.4**
218	Imi	-2.97	20	15	17	15	**12.1**
160	Imi	-4.01	22	20	**13.0**	5	9
211	Pla	-2.13	21	**16.5**	14	19	**15.1**
133	Imi	-5.66	20	10	7	2	**2.7**
206	Imi	-5.97	22	9	3	**2.8**	1
233	Pla	-5.43	20	11	4	**4.1**	0
114	Cbt	-4.79	21	8	8	**7.3**	**6.7**
226	Ipt	-4.87	25	10	**8.2**	**6.9**	**5.6**
161	Imi	-5.91	22	5	**4.5**	**4.1**	**2.5**

2.1 *A Three-Level model.* The previously described mixed-effects model can be expanded to include two levels of nesting. To express the 3-level model in a general way, it is useful to use a matrix representation of the model. Stacking the response vectors of each subject within a 3-level unit, the 3-level model for the resulting N_i response vector for the ith 3-level unit (classroom, clinic, *etc.*,), $i = 1, 2, \ldots, N$, can be written as follows:

$$
\underset{N_i \times 1}{\begin{bmatrix} \boldsymbol{y}_{i\,1} \\ \boldsymbol{y}_{i\,2} \\ \boldsymbol{y}_{i\,3} \\ \cdots \\ \cdots \\ \boldsymbol{y}_{i\,n_i} \end{bmatrix}}
= \underset{N_i \times ((n_i \times r)+1)}{\begin{bmatrix} \mathbf{1}_{i\,1} & \boldsymbol{X}_{i\,1} & 0 & 0 & \cdots & 0 \\ \mathbf{1}_{i\,2} & 0 & \boldsymbol{X}_{i\,2} & 0 & \cdots & 0 \\ \mathbf{1}_{i\,3} & 0 & 0 & \boldsymbol{X}_{i\,3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{1}_{i\,n_i} & 0 & 0 & 0 & \cdots & \boldsymbol{X}_{i\,n_i} \end{bmatrix}}
\underset{((n_i \times r)+1) \times 1}{\begin{bmatrix} \beta_{0\,i} \\ \boldsymbol{\beta}_{i1} \\ \boldsymbol{\beta}_{i2} \\ \boldsymbol{\beta}_{i3} \\ \cdots \\ \boldsymbol{\beta}_{i n_i} \end{bmatrix}}
$$

$$
+ \underset{N_i \times p}{\begin{bmatrix} \mathbf{1}_{i\,1} & \boldsymbol{W}_{i\,1} \\ \mathbf{1}_{i\,2} & \boldsymbol{W}_{i\,2} \\ \mathbf{1}_{i\,3} & \boldsymbol{W}_{i\,3} \\ \cdots & \cdots \\ \cdots & \cdots \\ \mathbf{1}_{i\,n_i} & \boldsymbol{W}_{i\,n_i} \end{bmatrix}}
\underset{p \times 1}{\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \cdots \\ \alpha_p \end{bmatrix}}
+ \underset{N_i \times 1}{\begin{bmatrix} \boldsymbol{\varepsilon}_{i\,1} \\ \boldsymbol{\varepsilon}_{i\,2} \\ \boldsymbol{\varepsilon}_{i\,3} \\ \cdots \\ \cdots \\ \boldsymbol{\varepsilon}_{i\,n_i} \end{bmatrix}} \tag{5}
$$

where, $\beta_{0\,i} \sim \mathcal{N}(0, \sigma^2_{\beta(3)})$, $\boldsymbol{\beta}_{ij} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\beta(2)})$, and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2_{\varepsilon} \boldsymbol{\Omega}_i)$. Notice, there are n_i subjects within site i and N_i total observations within site i (the sum of all repeated observations for all subjects within the site). The number of random subject-level effects is r and the number of fixed covariates in the model (including the intercept) is p. Each person has a $n_{ij} \times 1$ vector $\boldsymbol{y}_{ij}$ of repeated observations of the dependent variable, a $n_{ij} \times r$ design matrix $\boldsymbol{X}_{ij}$ for their r random effects $\boldsymbol{\beta}_{ij}$, and a $n_{ij} \times p$ matrix of covariates $\boldsymbol{W}_{ij}$. The covariate matrix usually includes the mixed-effects design matrix so that the overall intercept, linear term, *etc.*, is estimated and thus the random effects represent deviations from these overall terms. In terms of the autocorrelated errors, although $\boldsymbol{\Omega}_i$ carries the i subscript, it depends on i only through its dimension N_i, that is, the number of parameters in $\boldsymbol{\Omega}_i$ will not depend on i. Note that within a three-level cluster, the residuals are not correlated between individuals, thus $\boldsymbol{\Omega}_i$ has the form:

$$
\boldsymbol{\Omega}_i = \begin{bmatrix} \boldsymbol{\Omega}_{i\,1} & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Omega}_{i\,2} & 0 & \cdots & 0 \\ 0 & 0 & \boldsymbol{\Omega}_{i\,3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{\Omega}_{i\,n_i} \end{bmatrix} .
$$

With these assumptions, the observations $\boldsymbol{y}_i$ and random coefficients $\boldsymbol{\beta}$ have a joint multivariate normal distribution as in equation (2), where the distribution of the random coefficients $\boldsymbol{\beta}_i$, is:

$$\begin{bmatrix} \beta_{0\,i} \\ \boldsymbol{\beta}_{i1} \\ \boldsymbol{\beta}_{i2} \\ \cdots \\ \boldsymbol{\beta}_{in_i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \begin{bmatrix} \sigma^2_{\beta(3)} & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_{\beta(2)} & 0 & \cdots & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{\beta(2)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{\Sigma}_{\beta(2)} \end{bmatrix} \right).$$

Parameter estimation is a direct extension of the two-level case as described in the Appendix.

2.1.1. *Illustration.* In the previous illustration, we considered a two-level model for longitudinal response data from the NIMH TDCRP study. In these analyses, there are clearly significant patient related effects that predispose the rate of change over time to various treatments. We note however, that subjects are nested within therapists and both observable (*e.g.*, psychiatrist versus psychologist or numbers of years of experience) and unobservable therapist characteristics may play a role in the efficacy of the given treatment. To this end, we applied a three-level extension of the original model described by Gibbons and co-workers (1993) in which both random patient and therapist effects were jointly estimated. In addition, we included covariates at the level of the therapist that included years of experience and discipline (*i.e.*, psychiatrist versus psychologist). At the level of the patient, covariates included week, and treatment (*i.e.*, CBT versus IPT since therapists are only relevant to the two psychotherapy conditions) and the week by treatment interaction.

Results of our reanalysis of these data revealed several interesting effects. First, the therapist effect was not statistically significant. In the scale of 17-item HRSD scores, the random effect standard deviations were .53 for therapist, 2.36 for patient trends over time (*i.e.*, HRSD versus natural logarithm of week plus 1.0) and 3.82 for residual variation. These results yield intra-class correlations of .02 for therapist and .27 for patient. As such, therapist to therapist variability does not play a significant role in the variance decomposition once the effects of treatment, time, therapist experience and discipline are accounted for.

In terms of the measurable therapist effects of discipline and experience, no significant discipline related effects were observed. In terms of experience, the patterns of change over time were independent of years of experience of the therapists, however, a significant experience by treatment interaction ($p < .004$) was found indicating that averaging over time, the HRSD scores were differentially related to therapist experience in the two treatment groups (*i.e.*, CBT versus IPT). The source of this interaction appears to be an overall decrease in HRSD scores (*i.e.*, less severe) with therapist experience for CBT whereas an overall increase in HRSD scores with therapist experience for IPT. Inspection of the observed means over time reveal that the interaction is largely due to a larger week 4 decrease in HRSD scores for CBT patients with more experienced therapists. The effect for IPT was largely due to the more experienced therapists having more severely depressed patients at baseline. These effects are only significant averaging over time and therapist characteristics played no role in changing the overall rate of change over time.

In terms of patient level fixed effects, there were no significant treatment related differences between the two psychotherapies (*i.e.*, main effect of treatment or treatment by time interaction).

Figure 3 presents a graphical display of the empirical Bayes estimates of week 16 HRSD scores for individual subjects (obtained from the three-level model individual patient slope estimates) and average therapist week 16 HRSD estimates (obtained from the random therapist effect and mean intercept and slope of the regression). The average therapist effects are the stars connected by a line in the center of the plot and the squares represent individual patient week 16 HRSD estimates for each therapist. Inspection of the plot reveals that a small subgroup of CBT patients have high estimates of week 16 HRSD scores which is consistent with previous observations that IPT performed somewhat better than CBT (see Gibbons *et al.*, 1993). Interestingly, the IPT psychiatrist (denoted MD on the graph) and psychologist with the best outcomes had the least amount of experience (*i.e.*, therapist 9 (a psychiatrist) with 3 years of experience and therapist 15 (a psychologist) with 12 years of experience respectively). By contrast, the therapist with best results for CBT (*i.e.*, therapist 8) had the most experience. This observation is consistent with the previously described treatment by experience interaction. There do not appear to be any consistent differences between psychologists and psychiatrists in estimated week 16 HRSD scores. As is obvious by the overlap in Figure 3, none of the differences between treatment, profession or years of experience significantly effect outcome.

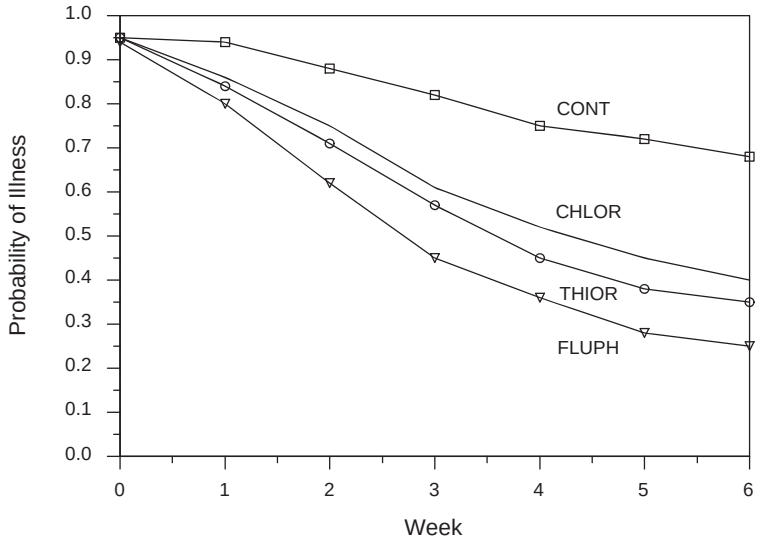

Figure 3. Empirical Bayes estimates of Week 16 HRSD score. Therapist and patient estimates sorted by years of experience.

2.2 *Modeling nonignorable nonresponse.* As presented to this point, the mixed-effects model assumes that the missing data are "ignorable" conditional on both

the covariates in the model and the available responses for that subject (Laird, 1988). This is an important distinction between full-likelihood approaches as described in the previous section and alternative partial-likelihood procedures, such as generalized estimating equations (GEE), Liang and Zeger (1986) and Zeger and Liang (1986), which assume no specific distributional form for the response measure but are less flexible in terms of their assumption regarding missing data. GEE type models for longitudinal data assume that the missing data are ignorable conditional on the covariates alone. Information contained in the available data for that subject prior to dropout do not provide a basis for ignorability of the missing data as they do in the full-likelihood approach.

In certain cases, the missing data are not related to prior responses or covariates in the model; therefore they are "nonignorable" even in the full-likelihood case. Recently, Little (1993, 1994 and 1995) has described a general class of models dealing with missing data under the rubric of "pattern-mixture models." In particular, Little (1995) and Hedeker and Gibbons (1997) have presented one solution to this problem based on mixed-effects pattern-mixture models for longitudinal data with drop-outs in which the usual ignorable missing data assumption is too restrictive. In these models, subjects are divided into groups depending on their missing-data pattern. These groups then can be used, for example, to examine the effect of the missing-data pattern on the outcome(s) of interest. For example, suppose that subjects are measured at three time-points; then there are $2^3 = 8$ possible missing data patterns. By grouping subjects in this way, or in reasonable subsets of possible missing data patterns (including the simplest subset *i.e.*, completer versus dropout) we have created a between-subjects variable, the missing data pattern, that can be included in a mixed-effects model as a covariate. Of course, this will only work for models that allow missing data since models that require complete data would have missing data for the missing data pattern covariate. Based on the number of missing data patterns selected, dummy coded variables representing deviations from the nonmissing pattern can be entered into the mixed-effects model. In this way, one can examine (a) the degree to which the missing data patterns differ in terms of outcome and (b) the degree to which the missing data pattern moderates the other fixed effects in the model (*e.g.*, treatment). Hedeker and Gibbons (1997) go on to illustrate how submodels can be obtained for each missing data pattern and how to obtain overall averaged estimates and standard errors for each treatment (*i.e.*, averaging over missing data patterns). Results of an application of this method are presented in a following section.

Besides the pattern-mixture approach, other methods have been proposed to handle missing data in longitudinal studies (Heckman, 1976; Diggle and Kenward, 1994). These alternative approaches are termed "selection models," and involve two stages which are either performed separately or iteratively. The first stage is to develop a predictive model for whether or not a subject drops out, using variables obtained prior to the dropout, often the variables measured at baseline. This model of dropout provides a predicted dropout probability or propensity for each subject; these dropout propensity scores are then used in the (second stage) longitudinal data model as a covariate to adjust for the potential influence of dropout. While

the selection models provide valuable information on what the predictors of study dropout might be, an advantage of the pattern-mixture models is that they can be used even when no such predictors are available. Also, some authors (Little and Rubin, 1987) have pointed out that the adjustment provided by selection models largely rests on assumptions which are difficult to empirically assess (*e.g.*, do you have the correct predictors of drop-out in the first place).

3. Models for Binary Data

While there has been considerable interest in mixed-effects models for longitudinal and hierarchical, clustered, or multi-level measurement data, there has been less focus on mixed-effects models for discrete data. Stiratelli *et al.*, (1984) developed a mixed-effects logit model for modeling correlated binary data and Gibbons and Bock (1987) developed a more general mixed-effects probit model for similar applications. Gibbons *et al.*, (1994) and Gibbons and Hedeker (1994) further generalized the mixed-effects probit model for application to multiple time-varying and time-invariant covariates and alternate response functions and prior distributions. Using quasi-likelihood methods in which no distributional form is assumed for the outcome measure, Liang and Zeger (1986) and Zeger and Liang (1986) have shown that consistent estimates of regression parameters and variance estimates can be obtained regardless of time dependence. Koch *et al.*, (1977) and Goldstein (1991) have illustrated how random effects can be incorporated into log-linear models. Generalizations of the logistic regression model in which the values of all regression coefficients vary randomly over individuals have been proposed by Wong and Mason (1985) and Conoway (1989). These models are applicable to both longitudinal and clustered problems; many are described in the recent review articles of Fitzmaurice, Laird, & Rotnitzky (1993) and Pendergast *et al.*, (1996). In the following, some details of the mixed-effects probit/logistic regression models are provided.

3.1 *Mixed-effects models for longitudinal binary data.* The general theory introduced in the previous section can be used to derive a mixed-effects probit or logistic regression model for the analysis of longitudinal binary data. In the case of a binary outcome and a single random effect (*e.g.*, a random intercept model) the general mixed-effects regression model in (1) now describes $\boldsymbol{y}_i$ as a vector of unobservable continuous "response strengths", for example:

$$y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \alpha_2\, x_{2i} + \alpha_3\, x_{3ij} + \beta_{0i} + \varepsilon_{ij}\ , \tag{6}$$

where

y_{ij}	=	the unobservable continuous "response strength" or "propensity" on time-point j for subject i
t_{ij}	=	is the time (*i.e.*, day, week, year etc.) that corresponds to the jth measurement for subject i
β_{0i}	=	the random effect for subject i (*i.e.*, deviation from the population intercept α_0.
α_2	=	the fixed effect of the subject level covariate x_{2_i}
α_3	=	the fixed effect of the time-specific covariate $x_{3_{ij}}$
ε_{ij}	=	an independent residual distributed $\mathcal{N}(0, \sigma_\varepsilon^2)$ for a probit model or assuming a logistic distribution with variance $\sigma_\varepsilon^2 \pi^2/3$ for a logistic regression model.

If we assume that distribution of the β_i is normal $N(0, \sigma_\beta)$, then the conditional probability for the binary response pattern of subject i (*i.e.*, $\boldsymbol{Y}_i$) is

$$\ell(\boldsymbol{Y}_i \mid \beta; \boldsymbol{\alpha}) = \prod_{j=1}^{n_i} [\Phi(z_{ij})]^{Y_{ij}} [1 - \Phi(z_{ij})]^{1-Y_{ij}} \ , \tag{7}$$

where

$$z_{ij} = (\beta_i + \alpha_0 + \alpha_1 t_{ij} + \alpha_2 \, x_{2i} + \alpha_3 \, x_{3ij} - \gamma)/\sigma_\varepsilon.$$

Here, γ represents a threshold on the underlying distribution and without loss generality we can let $\sigma_\varepsilon = 1$ and $\gamma = 0$. Thus, the marginal probability is:

$$h(\boldsymbol{Y}_i) = \int_\beta \ell(\boldsymbol{Y}_i \mid \beta; \boldsymbol{\alpha}) \, g(\beta) \, d\beta, \tag{8}$$

where $g(\beta)$ is the normal density with mean 0 and variance σ_β^2. Alternatively, we can express β_i in standardized form *i.e.*,

$$z_{ij} = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 \, x_{2i} + \alpha_3 \, x_{3ij} + \sigma_\beta \theta_i \ ,$$

where $\theta_i \ \sim \ N(0, 1)$. In the general case of multiple random effects, Gibbons and Bock (1987) orthogonally transform the response model such that $\boldsymbol{\beta} = \boldsymbol{\Lambda\theta}$, where $\boldsymbol{\Lambda\Lambda}' = \boldsymbol{\Sigma}_\beta$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_\beta$. A consequence of transforming from $\boldsymbol{\beta}$ to $\boldsymbol{\theta}$ is that the Cholesky factor $\boldsymbol{\Lambda}$, which is a lower triangular matrix, is estimated instead of the covariance matrix $\boldsymbol{\Sigma}_\beta$. As the Cholesky factor is essentially the square-root of the covariance matrix, this then allows more stable estimation of near-zero variance terms.

The general method of estimation for the parameters of this nonlinear model was originally described by Gibbons and Bock (1987). Gibbons and Hedeker (1997) have introduced a three-level version of this model.

3.1.1 *Illustration.* Gibbons and Hedeker (1994) illustrated application of mixed-effects probit models in a reanalysis of the National Institute of Mental Health Schizophrenia Collaborative Study. The study was designed to compare three anti-psychotic medications to placebo in a large randomized clinical trial. The study

is of historical importance because it represents one of the last studies in which schizophrenics were allowed to be randomized to a placebo. In this analysis, the binary outcome was overall severity of illness, where 0 represents not ill to mildly ill and 1 represents moderately ill to severely ill. Hedeker and Gibbons (1994) performed a similar analysis treating the responses on an ordinal scale. Experimental design and corresponding sample sizes are displayed in Table 5.

Table 5. Experimental Design and Weekly Sample Sizes

Treatment	Sample Size at Week						
Group	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 5 reveals that the longitudinal portion of the study is highly unbalanced with large differences in the number of measurements made in the six weeks of treatment. Results of fitting a fixed-effects model and mixed-effects models with one and two random effects are presented in Table 6. Table 6 reveals different results depending on whether or not random effects are included. The model with one random effect significantly improved fit over the fixed-effects model ($\chi^2_1 = 131.04,\ p < .0001$), and the model with two-random effects significantly improved fit over the model with one random effect ($\chi^2_1 = 73.70,\ p < .0001$). Person-specific variability in intercepts $\hat{\sigma}_{\beta_0} = .860$ and slopes $\hat{\sigma}_{\beta_1} = .630$ were both significant ($p < .0001$), but uncorrelated ($\hat{\sigma}_{\beta_0\beta_1} = .056, p < .48$).

Table 6. Parameter Estimates, Standard Errors and Probabilities for NIMH Schizophrenia Collaborative Study

	Fixed			1 Mixed			2 Mixed		
Fixed Effects	MLE	SE	P <	MLE	SE	P <	MLE	SE	P <
Intercept	-1.777	.158	.0001	-2.630	.314	.0001	-2.507	.457	.0001
Slope	.217	.037	.0001	.309	.042	.0001	.102	.105	.331
Sex	.126	.056	.02	.178	.140	.20	.215	.188	.25
Chlor vs Pla	.265	.175	.13	.395	.285	.16	.050	.285	.86
Fluph vs Pla	.516	.187	.006	.810	.303	.008	.209	.339	.54
Thior vs Pla	.314	.170	.07	.357	.284	.21	.079	.284	.78
C vs Pla by T	.064	.050	.20	.111	.055	.04	.427	.136	.002
F vs Pla by T	.102	.054	.06	.164	.061	.007	.706	.155	.0001
T vs Pla by T	.078	.050	.12	.165	.061	.007	.526	.144	.0002
Random Effects									
σ_{β_0}				1.180	.112	.0001	.860	.220	.0001
$\sigma_{\beta_0\alpha_1}$							.056	.093	.48
σ_{β_1}							.630	.112	.0001
Log L	-780.81			-715.29			-678.44		
Change χ^2				131.04			73.70		
Change *df*				1			2		
Change $p <$				.0001			.0001		

Sex, main effect of treatment and treatment by time interaction were examined. Both mixed-effects models revealed significant treatment by time interactions for all

three active treatments versus placebo control, although magnitude was somewhat greater for the model with two random effects indicating that differences between treatment groups and the placebo control group were linearly increased over the six week study. However, the fixed-effects model did not identify significant treatment by time interactions. Only the main effect (*i.e.*, averaging over time-points) corresponding to difference between Fluphenazine and placebo was significant. In contrast, the fixed-effects model identified a significant sex effect not found in either mixed-effects model.

These results illustrate that ignoring systematic person-specific effects leads to poor model fit, and can bias the maximum likelihood estimates, standard errors, and probability values associated with tests of treatment-related effects. Indeed, had we naively applied a traditional probit or logistic regression model to these data, we would have incorrectly concluded Thioridazine and Chlorpromazine did not have any beneficial effects relative to placebo control.

To better understand these differences, predicted probability of illness curves are displayed in Figure 4. The treatment versus control differences are clearly evident in the Figure, with consistent differences emerging as early as one week.

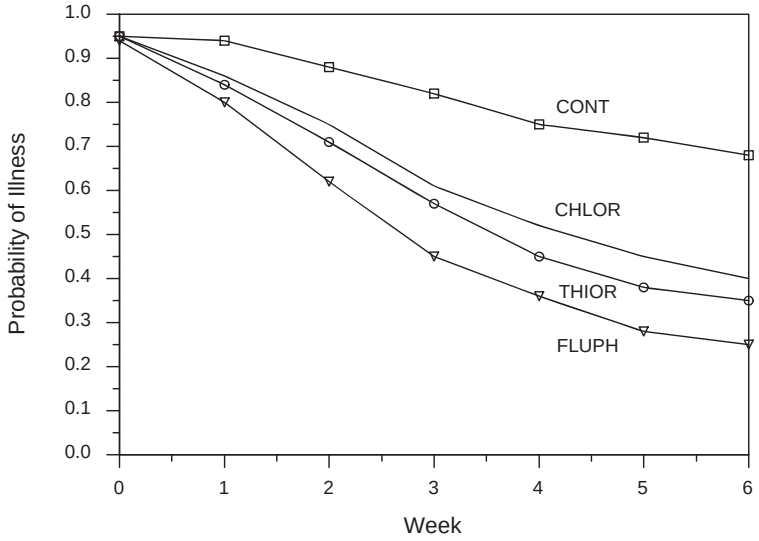

Figure 4. Estimated probability of illness curves by treatment group. CONT = control, CHLOR = chlorpromazine, THIOR = thioridazine, FLUPH = fluphenazine.

Hedeker and Gibbons (1997) applied a pattern-mixture model to these data to determine the extent to which the missing data could be considered ignorable. Although they found that there was a significant drug by time by dropout interaction, in which the combined drug versus control difference was more pronounced in those subjects who did not complete the study, the overall drug effect and drug by time interaction were virtually identical to the mixed-effects model which assumed ignorable nonresponse.

3.2 *Multivariate mixed-effects probit models.* Probit analysis (Bliss, 1935) is classically used to describe a dosage response relation between a continuous dosage metameter and a quantal (*i.e.*, presence or absence, alive or dead) response. In multivariate probit analysis (Ashford and Sowden, 1970), a continuous grouping variable (*e.g.*, dosage) is assumed to statistically control the joint response on n-quantal variables. The n-quantal response variables may describe the presence or absence of an abnormality in multiple biological systems, the joint occurrence of several psychopathological symptoms in a single patient, a set of coexisting attitudes measured by a sociological survey, a pattern of correct or incorrect responses on a test or a series of discrete choices made by a consumer. These n-quantal variables may represent a multivariate observation at a single point in time or a series of univariate observations made repeatedly over time in the same individual.

Ashford and Sowden (1970) generalized probit analysis to the bivariate case by assuming that the joint response of two biological systems followed a bivariate normal distribution with correlation coefficient r. This generalization provides a full likelihood solution for the bivariate case, but is computationally intractable for more than two quantal variates. Muthén (1979) introduced a more general probit model for n dichotomous indicators of m continuous latent variables. Muthén proposed a psychometric measurement model (Bock and Lieberman 1970; Lord and Novick 1968) to relate a p-dimensional vector of dichotomous response variables to an m dimensional vector of continuous latent variables. Structural relations (*e.g.*, dose responses) are then estimated in terms of the m latent variables and not n manifest quantal responses.

Bock and Gibbons (1996), Gibbons and Wilcox-Gök, (1998), Gibbons and Lavigne (1998) generalized Ashford and Sowden's (1970) results as a multivariate mixed-effects probit model which can be applied simultaneously to a large number of quantal response variables. They assume that each subject has a covariate vector $\boldsymbol{W}_i$ of length p that can be any mixture of discrete and continuous variables. Each subject produces n distinct quantal responses or is classified with respect to n dichotomous variables. The quantal responses for subject i are accounted for by a n-vector of underlying "response strengths"

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{A} + \boldsymbol{\Lambda}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i \ ,$$

where $\boldsymbol{W}_i$ is the p-vector of covariates associated with subject i and $\boldsymbol{A}$ is a n x p matrix of coefficients of the regression of $\boldsymbol{y}$ on $\boldsymbol{W}$. The r-vector θ, $r < n$, contains values of the underlying factors (*i.e.*, random effects or latent variables) that account for correlation among the n quantal response variables through the n x r matrix of factor coefficients, $\boldsymbol{\Lambda}$. Ultimately, these correlations are expressed in the associations among the quantal variables. Finally, $\boldsymbol{\varepsilon}$ is a n-vector of uncorrelated residuals.

Under multivariate normal assumptions there is no loss of generality in assuming

$$\boldsymbol{\theta} \sim g(\boldsymbol{\theta}) \sim N(0, \boldsymbol{I}) \quad \text{and} \quad \boldsymbol{\varepsilon} \sim N(0, d_j^2 \boldsymbol{I}_n) \ ,$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are mutually independent and d_j^2 is the unique variance for quantal response variate j (*i.e.*, $d_j^2 = 1 - \sum_{q=1}^{r} \lambda_{jq}^2$).

Joint distribution of $\boldsymbol{y}_i$ and $\boldsymbol{\theta}$ is therefore $(p+m)$-variate normal:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{\theta} \end{bmatrix} = N\left(\begin{bmatrix} \boldsymbol{W}_i\boldsymbol{A} \\ \boldsymbol{0} \end{bmatrix} , \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + d_j^2\boldsymbol{I}_n & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}' & \boldsymbol{I} \end{bmatrix} \right) .$$

Bock and Aitkin (1981) develop the factor portion of the model by assuming that for each binary response variable a positive response by subject i on variable j (*i.e.*, $Y_{ij} = 1$) occurs when the corresponding underlying response strength (y_{ij}) exceeds a threshold γ_j; otherwise, $Y_{ij} = 0$. Thus on the previous assumptions regarding normality and independence of the elements of $\boldsymbol{\varepsilon}$, the probability of a positive response for subject i on variable j, conditional on $\boldsymbol{\theta}$, is

$$\begin{aligned} P(Y_{ij} = 1|\boldsymbol{\theta}_i) &= \frac{1}{\sqrt{2\pi d_j^2}} \int_{\gamma_j}^{\infty} \exp\left\{ -\frac{1}{2} \left[y_{ij} - (\boldsymbol{w}_{ij}\boldsymbol{\alpha}_j + \boldsymbol{\lambda}_j\boldsymbol{\theta}_i)/d_j \right]^2 \right\} dy \\ &= \Phi(-(\gamma_j - z_{ij})/d_j) , \end{aligned} \tag{9}$$

where Φ is the standard univariate normal distribution function and

$$z_{ij} = \boldsymbol{w}_{ij}\boldsymbol{\alpha}_j + \boldsymbol{\lambda}_j\boldsymbol{\theta}_i ,$$

and $\boldsymbol{\alpha}_j$ and $\boldsymbol{\lambda}_j$ are the j-th rows of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$, respectively. Without loss of generality, the origin and unit can be chosen arbitrarily. For convenience, let $\gamma_j = 0$ and $d_j = 1$.

On these assumptions, the quantal responses are conditionally independent given $\boldsymbol{\theta}$, and the conditional probability of observing n-vector $\boldsymbol{Y}_i = [Y_{ij}]$ is

$$L_i(\boldsymbol{\theta}) = \prod_j^n \left[\Phi(z_{ij}) \right]^{Y_{ij}} \left[1 - \Phi(z_{ij}) \right]^{1-Y_{ij}} .$$

Moreover, because the components of $\boldsymbol{\theta}$ are uncorrelated, the marginal probability of observing the pattern $\boldsymbol{Y}_i = [1, 1, \ldots, 1]$ is

$$P(\boldsymbol{Y}_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_i(\boldsymbol{\theta}) \prod_q^r g(\theta_q) d\theta_1 d\theta_2 \cdots d\theta_r ,$$

where $g(\theta_q)$ is the standard normal ordinate at θ_q *i.e.*, $\phi(\theta_q)$. This definite integral is the probability of positive orthant of the n-variate distribution for a subject with covariate matrix $\boldsymbol{W}_i$. To obtain any other orthant probability, we reverse direction of integration for those variables with $Y_{ij} = 0$. Evaluation of these integrals and parameter estimation are closely related to the mixed-effects binary probit model and are described in detail by Bock and Gibbons (1996).

3.2.1 *Illustration.* Ashford and Sowden (1970) illustrated their method using two of five symptoms collected as part of the National Coal Board's Pneumoconiosis Field Research Project (Fay, 1957). These data were obtained by periodic medical examinations of working coal miners in the United Kingdom. Subjects were

asked to report on five respiratory symptoms; excessive coughing, phlegm, breathlessness, wheeze and whether or not their symptoms were dependent on the weather. The data set consists of 18,000 subjects classified as smokers without radiological pneumoconiosis between the ages of 20 and 64 years of age.

Bock and Gibbons (1996) have presented the 32 response pattern frequencies for the full five symptom data separately for each age group. Based on the complete five-symptom data, factor loadings, trend parameter estimates and standard errors are presented in Table 7.

Table 7. Factor loadings, trend parameter estimates and standard errors

	$\hat{\lambda}_{j1}$	$\hat{\lambda}_{j2}$	$\hat{\lambda}_{j3}$	$\hat{\lambda}_{j4}$	$\hat{\beta}_{j0}$	$\hat{\beta}_{j1}$	$\hat{\beta}_{j2}$
Cough	0.5905	0.6564	0.1709	0.3082	−2.0532	1.3647	−0.1069
Phlegm	0.6081	0.6209	0.1912	0.3110	−2.2032	1.1684	−0.0245
Breathl	0.1732	0.4314	0.5148	0.5203	−3.6965	2.2659	−0.0518
Wheeze	0.1837	0.4433	0.4552	0.6469	−2.7049	1.5981	−0.0848
Weather	−0.0717	0.6519	0.3256	0.6347	−2.7424	1.6240	−0.0900
	0.00000	0.00000	0.00000	0.00000	0.02473	0.02553	0.02435
Standard	0.01183	0.00000	0.00000	0.00000	0.02454	0.02514	0.02434
Errors	0.03988	0.03930	0.00000	0.00000	0.03624	0.04019	0.03327
	0.04076	0.03698	0.02089	0.00000	0.02727	0.02879	0.02644
	0.04041	0.03988	0.03243	0.02026	0.02745	0.02894	0.02663

Results of the analysis reveal significant linear age-related trends for all five symptoms. In contrast to the original analysis, there is also evidence of significant nonlinearity in that the quadratic orthogonal polynomial term was significant. The five symptoms required two factors to properly model the inter-symptom correlation matrix which yielded correlations in the range of .60 to .85 indicating strong association among the five symptoms (see Table 8).

Table 8. Correlation matrix and standard errors

	Cough	Phlegm	Breathl	Wheeze	Weather
Cough	1.0000				
Phlegm	0.8952	1.0000			
Breathl	0.6337	0.6334	1.0000		
Wheeze	0.6766	0.6752	0.7939	1.0000	
Weather	0.6369	0.6208	0.7666	0.8346	1.0000
	0.00606				
Standard	0.01002	0.01029			
Errors	0.00915	0.00973	0.01272		
	0.00973	0.00971	0.01320	0.01347	

The estimated factor loadings revealed that coughing and phlegm characterized the first factor and breathlessness, wheeze and weather sensitivity characterized the second factor. The observed and estimated age trend lines for each symptom are displayed graphically in Figure 5.

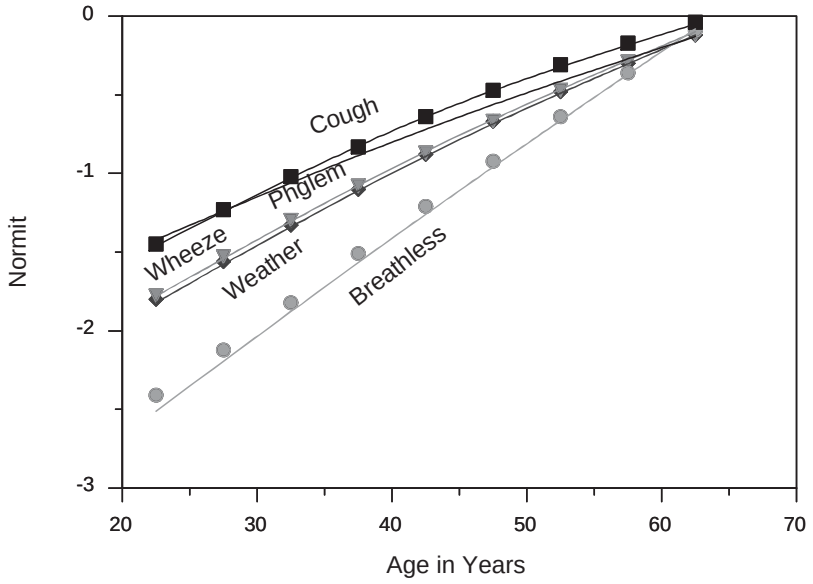

Figure 5. Observed and expected prevalence of five respiratory symptoms as a function of age. Prevalence expressed as a point on the normal distribution (i.e., normit scale)

4. Models for Ordinal Response Data

Ordinal response data are common in biomedical studies, for example, subjects may be classified in terms of exhibiting definite, mild, or no symptomatology of a given disease or condition. For ordinal responses, several authors have described models including both random and fixed effects. Harville and Mee (1984) and Jansen (1990) both describe ordinal probit models implementing the EM algorithm for parameter estimation. Ezzet and Whitehead (1991) provide a random-intercepts proportional odds model for a crossover trial using the Newton-Raphson method. A random-intercepts proportional odds model is also described by Agresti and Lang (1993), however, their approach is limited to within-cluster covariates. Utilizing the complementary log-log link function, Ten Have (1996) describes a random-intercepts model for ordinal responses assuming a log-gamma random effects distribution. Hedeker and Gibbons (1994) describe both an ordinal probit and logistic model with multiple random effects, and allow for both within and between-cluster covariates.

All of the above models utilizing the ordinal logistic regression formulation include the proportional odds assumption (McCullagh, 1980) for model covariates. This assumption implies that the effect of a regressor variable is the same across the cumulative logits of the model, or proportional across the cumulative odds. As noted by Peterson and Harrell (1990), however, examples of non-proportional odds are not difficult to find. Recently, Hedeker and Mermelstein (1998) have described an extension of the random effects proportional odds model to allow for

non-proportional odds for a set of explanatory variables. The resulting mixed-effects partial proportional odds model follows Peterson and Harrell's (1990) extension of the fixed-effects proportional odds model. Here, we will describe both the proportional and partial proportional odds models, and illustrate application of these models.

4.1 *Mixed-effects proportional odds model.* Assume that there are $i = 1, \ldots, N$ level-2 units and $j = 1, \ldots, n_i$ level-1 units nested within each level-2 unit. The cumulative probabilities for the k ordered categories ($k = 1, \ldots K$) are defined for the ordinal outcomes Y as:

$$P_{ijk} = \Pr\ (Y \leq k \mid \boldsymbol{\beta}_i; \gamma_k, \boldsymbol{\alpha})\ , \tag{10}$$

where the mixed-effects logistic regression model for these cumulative probabilities is given as

$$\log \frac{P_{ijk}}{(1 - P_{ijk})} = \gamma_k + \boldsymbol{x}'_{ij}\boldsymbol{\beta}_i + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}\ , \tag{11}$$

with $K - 1$ strictly increasing model intercepts γ_k (*i.e.*, $\gamma_1 > \gamma_2 \ldots > \gamma_{K-1}$). As before, $\boldsymbol{w}_{ij}$ is the $p \times 1$ covariate vector and $\boldsymbol{x}_{ij}$ is the design vector for the r random effects, both vectors being for the jth level-1 unit nested within level-2 unit i. Also, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of unknown fixed regression parameters, and $\boldsymbol{\beta}_i$ is the $r \times 1$ vector of unknown random effects for the level-2 unit i. Since the regression coefficients $\boldsymbol{\alpha}$ do not depend on k, the model assumes that the relationship between the explanatory variables and the cumulative logits does not depend on k. McCullagh (1980) calls this assumption of identical odds ratios across the $K-1$ cutoffs the proportional odds assumption. Assumptions regarding the mixed-effects and missing or unbalanced data are identical to the previously described mixed-effects models. Again, it is convenient to orthogonally transform the response model. Letting $\boldsymbol{\beta} = \boldsymbol{\Lambda\theta}$, where $\boldsymbol{\Lambda\Lambda}' = \boldsymbol{\Sigma}_\beta$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_\beta$, the reparameterized model is then written as

$$\log \frac{P_{ijk}}{(1 - P_{ijk})} = \gamma_k + \boldsymbol{x}'_{ij}\boldsymbol{\Lambda\theta}_i + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}\ , \tag{12}$$

where $\boldsymbol{\theta}_i$ are distributed according to a multivariate standard normal.

As indicated earlier in the section on binary responses, in motivating probit and logistic regression models, it is often assumed that there is an unobservable latent variable (y) which is related to the actual response through the "threshold concept". For a binary response, one threshold value is assumed, while for an ordinal response, *a series* of threshold values $\gamma_1, \gamma_2, \ldots, \gamma_{K-1}$, where K equals the number of ordered categories, $\gamma_0 = -\infty$, and $\gamma_K = \infty$. Here, a response occurs in category k (Y = k) if the latent response process y exceeds the threshold value γ_{k-1}, but does not exceed the threshold value γ_k. As pointed out by McCullagh (1980), the threshold concept and its reference to the existence of a latent variable, though useful in motivating the ordinal model, is not required for model interpretation.

4.2 *Partial proportional odds.* To allow for a partial proportional odds model the intercepts γ_k are denoted instead as γ_{0k}, and the following terms are added to the model:

$$\log \frac{P_{ijk}}{(1 - P_{ijk})} = \gamma_{0k} + (\boldsymbol{u}^*_{ij})'\boldsymbol{\gamma}^*_k + \boldsymbol{x}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\theta}_i + \boldsymbol{w}'_{ij}\boldsymbol{\alpha} \tag{13}$$

or absorbing γ_{0k} and $\boldsymbol{\gamma}^*_k$ into $\boldsymbol{\gamma}_k$,

$$\log \frac{P_{ijk}}{(1 - P_{ijk})} = \boldsymbol{u}'_{ij}\boldsymbol{\gamma}_k + \boldsymbol{x}'_{ij}\boldsymbol{\Lambda}\boldsymbol{\theta}_i + \boldsymbol{w}'_{ij}\boldsymbol{\alpha} \tag{14}$$

where, $\boldsymbol{u}_{ij}$ is a $(h+1) \times 1$ vector containing (in addition to a 1 for γ_{0k}) the values of observation ij on the subset of h $\boldsymbol{w}_{ij}$ covariates for which proportional odds is not assumed. In this model, $\boldsymbol{\gamma}_k$ is a $(h+1) \times 1$ vector of regression coefficients associated with the h variables (plus the intercept) in $\boldsymbol{u}_{ij}$. Notice that the effects of these h covariates ($\boldsymbol{u}^*_{ij}$) vary across the $K-1$ cumulative logits. This extension of the model follows similar extensions of the ordinary fixed-effects ordinal logistic regression model discussed by Peterson and Harrell (1990) and Cox (1995). Terza (1985) discusses a similar extension for the ordinal probit regression model.

In general, this extension of the proportional odds model is not problematic, however, one caveat should be mentioned. For the explanatory variables without proportional odds, the effects on the cumulative log odds, namely $(\boldsymbol{u}^*_{ij})'\boldsymbol{\gamma}^*_k$, result in $K-1$ non-parallel regression lines. These regression lines inevitably cross for some values of $\boldsymbol{u}^*$, leading to negative fitted values for the response probabilities. For $\boldsymbol{u}^*$ variables contrasting two levels of an explanatory variable (*e.g.*, gender coded as 0 or 1), this crossing of regression lines occurs outside the range of admissible values (*i.e.*, < 0 or > 1). However, if the explanatory variable is continuous, this crossing can occur within the range of the data, and so, allowing for non-proportional odds is problematic. For continuous explanatory variables, other than requiring proportional odds, a solution to this dilemma is sometimes possible if the variable u has, say, m levels with a reasonable number of observations at each of these m levels. In this case $m-1$ dummy-coded variables can be created and substituted into the model in place of the continuous variable u.

With the above mixed-effects regression model, the probability of a response in category k for a given level-2 unit i, conditional on $\boldsymbol{\gamma}_k$, $\boldsymbol{\alpha}$, and $\boldsymbol{\theta}$ is given by:

$$\Pr(Y_j = k \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda}) = P_{j\,k} - P_{j\,k-1} \tag{15}$$

where, under the logistic response function,

$$P_{j\,k} = \frac{1}{1 + \exp(-z_{jk})}\,, \tag{16}$$

with $z_{jk} = \boldsymbol{u}'_j\boldsymbol{\gamma}_k + \boldsymbol{x}'_j\boldsymbol{\Lambda}\boldsymbol{\theta} + \boldsymbol{w}'_j\boldsymbol{\alpha}$. Note that $P_{j\,0} = 0$ and $P_{j\,k} = 1$.

Letting $\mathbf{Y}_{\boldsymbol{i}}$ denote the vector pattern of ordinal item responses from level-2 unit i for the n_i level-1 units nested within, the probability of any pattern $\mathbf{Y}_{\boldsymbol{i}}$, given $\boldsymbol{\gamma}_k$, and $\boldsymbol{\alpha}$, and $\boldsymbol{\theta}$ is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{Y_i} \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda}) = \prod_{j=1}^{n_i} \prod_{k=1}^{K} (P_{ijk} - P_{ijk-1})^{c_{ijk}} \tag{17}$$

$$\text{where } c_{ijk} = \begin{cases} 1 & \text{if } Y_{ij} = k \\ 0 & \text{if } Y_{ij} \neq k \end{cases} .$$

The marginal density of $\mathbf{Y_i}$ in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$:

$$h(\mathbf{Y_i}) = \int_{\boldsymbol{\theta}} \ell(\mathbf{Y_i} \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda})\, g(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{18}$$

where $g(\boldsymbol{\theta})$ represents the multivariate standard normal density. Estimation of the p covariate coefficients $\boldsymbol{\alpha}$, the population variance-covariance parameters $\boldsymbol{\Lambda}$ (with $r(r+1)/2$ elements), and the $(h+1)(K-1)$ parameters in $\boldsymbol{\gamma}_k$ $(k = 1, \ldots, K-1)$ is described in detail by Hedeker and Gibbons (1994) and summarized in the appendix.

Table 9. NORC DATA: OPINIONS ABOUT TEENAGE SEX, PREMARITAL SEX, AND EXTRAMARITAL SEX

Teen sex	Premarital sex	Extramarital sex 1	2	3	4
1	1	140	1	0	0
	2	30	3	1	0
	3	66	4	2	0
	4	83	15	10	1
2	1	3	1	0	0
	2	3	1	1	0
	3	15	8	0	0
	4	23	8	7	0
3	1	1	0	0	0
	2	0	0	0	0
	3	3	2	3	1
	4	13	4	6	0
4	1	0	0	0	0
	2	0	0	0	0
	3	0	0	1	0
	4	7	2	2	4

Note: 1=always wrong,
2=almost always wrong,
3=wrong only sometimes,
4=not wrong.

4.3 *Illustration.* Agresti and Lang (1993) previously examined the data presented in Table 9 using a cumulative logit model under the proportional odds assumption. These data are taken from the 1989 General Social Survey, conducted by the National Opinion Research Center at the University of Chicago. 475 subjects gave their opinion on three items: (1) early teens, age 14-16, having sex relations before marriage, (2) a man and a woman having sex relations before marriage, and (3) a married person having sexual relations with someone other than the marriage partner. These opinions were given in terms of four ordered categories: "always wrong," "almost always wrong," "wrong only sometimes," and "not wrong at all."

Table 10 lists results for analysis of these data considering both proportional and partial proportional odds models. In both models, a random subject effect is assumed to account for the dependency in the three repeated observations. Two contrasts (premarital versus teen sex, and extramarital versus teen sex) were also included as covariates in the models. In two models, the effects of these covariates were estimated assuming proportional and partial proportional odds, respectively.

Table 10. NORC data - Proportional and Partial Proportional Odds Models Maximum Marginal Likelihood Estimates (standard errors)

term	*Proportional Odds*		*Partial Proportional Odds*	
intercept 1 γ_{01}	2.047	***	1.808	***
	(.199)		(.191)	
intercept 2 γ_{02}	3.176	***	3.408	***
	(.250)		(.257)	
intercept 3 γ_{03}	4.812	***	5.026	***
	(.329)		(.390)	
Premarital sex α_1	-3.774	***		
	(.270)			
Premarital sex 1 γ_{11}			-3.152	***
			(.266)	
Premarital sex 2 γ_{12}			-4.124	***
			(.304)	
Premarital sex 3 γ_{13}			-4.219	***
			(.388)	
Extramarital sex α_2	.572	***		
	(.198)			
Extramarital sex 1 γ_{21}			.602	***
			(.199)	
Extramarital sex 2 γ_{22}			.335	
			(.281)	
Extramarital sex 3 γ_{23}			1.213	
			(.650)	
subject sd σ_β	2.234	***	2.106	***
	(.213)		(.195)	
$-2 \log L$	2437.84		2401.70	

$^{***}p < 0.01$ $^{**}p < 0.05$

Examining the results for the first model, we see significant effects for both covariates. Premarital sex is considered to be "not as wrong" (*i.e.*, opinions in the higher categories) as compared to teen sex. Conversely, extramarital sex is seen to be "more wrong" (*i.e.*, opinions in the lower categories) than teen sex. The variability attributable to subjects is highly significant and when expressed as an

intra-subject correlation equals .60, reflecting the high degree of dependency in these ordinal responses within subjects. The estimate of the intra-subject correlation is calculated according to the formula $\rho = \sigma_\beta^2/(\sigma_\beta^2 + \sigma_\varepsilon^2)$ where σ_ε^2 is the variance of the underlying continuous (standard logistic) latent variable y, discussed above, namely, $\pi^2/3$.

Comparing the second model to the first, the likelihood-ratio test clearly rejects the assumption of proportional odds for these two covariates considered jointly ($\chi^2 = 36.14, df = 4,\ p < .001$). Similarly, a multiparameter Wald test yields $\chi^2 = 34.96, df = 4,\ p < .001$ for a test of this assumption. Testing proportional odds for the two covariates separately yields Wald $\chi^2 = 22.54$ and 3.13 for premarital and extramarital sex, respectively (each on 2 df). Thus, based on these tests, it would appear that proportional odds may be reasonable for extramarital versus teen sex, but not for premarital versus teen sex.

To aid in interpretation, it is useful to calculate marginal empirical cumulative odds (and log odds) for each of the three items. These are calculated from Table 9 as 2.99, 8.69, and 28.69 (log odds 1.10, 2.16, and 3.36) for teen sex; .44, .64, and 1.57 (log odds -.81, -.45, and .45) for premarital sex; and 4.40, 11.18, and 78.17 (log odds 1.48, 2.41, and 4.36) for extramarital sex. In agreement with the parameter estimates in Table 10, the observed difference in log odds between teen sex and premarital sex is more pronounced for the latter two (-2.61 and -2.91, respectively) than for the first comparison (-1.91), indicating the non-proportional odds relationship. Conversely, comparing extramarital and teen sex yields observed log odds differences that are relatively similar (.38, .25, and 1.00, respectively) and rely on sparse data (6 and 16 of 475 subjects gave a response in the 4th category for teen and extramarital sex, respectively). Note that, as discussed by Neuhaus, Kalbfleisch, and Hauck (1991), to the degree that there is positive intra-cluster correlation, the mixed-effects model produces (cluster-specific) log-odds estimates that are larger in absolute value than the observed marginal log-odds.

5. Computer Programs

Computer software for estimating mixed-effects models is becoming increasingly available, especially for normal-theory models (MLn, Rasbash, Yang, M., Woodhouse, G., & Goldstein, 1995; HLM, Bryk, Raudenbush, & Congdon, 1996; VARCL, Longford, 1986; the BMDP 5V procedure, Schluchter, 1988; the SAS procedure MIXED). A detailed comparison of some of these programs is included in Kreft, de Leeuw, and van der Leeden (1994). Also, the multilevel homepage (http://www.uic.edu/multilevel/) provides a wealth of information about the MLn program in particular, and mixed-effects analysis in general. For the analyses of the NIMH TDCRP data presented in the current article, MIXREG (Hedeker and Gibbons, 1996b) was used. Software programs are also available for binary (EGRET, Statistics and Epidemiology Research Corporation, 1991) and ordinal (MIXOR, Hedeker and Gibbons, 1996a) response data. The MLn, HLM, and VARCL programs additionally have facilities for analysis of binary outcomes. In the current

article, MIXOR was used for the analyses presented in sections 3.1.1 and 4.3. Both MIXREG and MIXOR can be obtained free of charge through the internet from either the multilevel homepage or location http://www.uic.edu/~hedeker/mix.html.

6. Summary

Mixed-effects models have widespread application in biostatistics. As should be obvious from the review of the literature and various illustrations provided here, their potential application extends to many areas in the behavioral, social and physical sciences. The use of unobservable variables to account for correlation in clustered data is by no means a new one; however recent advances in computational statistics have made these models readily accessible to applied statisticians. Fully Bayesian approaches to the mixed-effects regression problem have also been considered (Johnson, 1996; Normand *et al.*, 1997; Albert and Chib, 1983). An advantage of the fully Bayesian approach is that it directly incorporates uncertainty in estimating the variance components which empirical Bayesian estimators ignore. Ignoring this uncertainty can lead to underestimates of variance components and uncertainty estimates of the fixed and random effects (Searle *et al.*, 1992; Seltzer, *et al.*, 1996). Interestingly, many biostatistical developments of mixed-effects regression models borrow strength from earlier work conducted in psychometrics (*e.g.*, Bock and Lieberman, 1970 and Bock and Aitkin, 1981) where person-specific effects are of considerable importance. There are many potentially fruitful areas for further statistical research. These areas include but are not limited to mixed-models for time-to-event data, counting processes and nominal outcomes (*e.g.*, discrete choice models). There are both clustered and longitudinal examples of each in biostatistical application as well as numerous other fields of interest.

Appendix

Maximum Marginal Likelihood Solution

Normally-distributed errors. The marginal density of the data $\boldsymbol{y}_i$ is given by

$$h(\boldsymbol{y}_i) = \int_{\boldsymbol{\beta}} f_i \cdot g \; d\boldsymbol{\beta} \; , \tag{19}$$

where, assuming normally-distributed errors,

$$f_i = (2\pi)^{-\frac{n_i}{2}} |\sigma_\varepsilon^2 \boldsymbol{\Omega}_i|^{-\frac{1}{2}} \; \exp\left[-\frac{1}{2} \boldsymbol{\varepsilon}_i' (\sigma_\varepsilon^2 \boldsymbol{\Omega}_i)^{-1} \boldsymbol{\varepsilon}_i \right] \; , \tag{20}$$

with $\boldsymbol{\varepsilon}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_i - \boldsymbol{W}_i \boldsymbol{\alpha}$. Here, the population distribution g of the random effects (sometimes called the prior distribution) is the multivariate normal density with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}_\beta$,

$$g = (2\pi)^{-\frac{r}{2}} |\boldsymbol{\Sigma}_\beta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}\right] . \tag{21}$$

The MML solution is then obtained by maximizing the log-marginal likelihood of the data from N level-2 units (*i.e.*, subjects in the longitudinal case),

$$\log L = \sum_{i=1}^{N} \log\left[h(\boldsymbol{y}_i)\right] \tag{22}$$

with respect to both the population parameters ($\boldsymbol{\Sigma}_\beta$) and the structural parameters ($\boldsymbol{\alpha}$, σ_ε^2, and $\boldsymbol{\omega}$). These equations are obtained as:

$$\hat{\boldsymbol{\Sigma}_\beta} = \frac{1}{N}\sum_{i}^{N} \tilde{\boldsymbol{\beta}}_i \tilde{\boldsymbol{\beta}}_i' + \boldsymbol{\Sigma}_{\beta|y_i} \tag{23}$$

$$\hat{\boldsymbol{\alpha}} = \left[\sum_{i=1}^{N} \boldsymbol{W}_i' \, \boldsymbol{\Omega}_i^{-1} \, \boldsymbol{W}_i\right]^{-1} \left[\sum_{i=1}^{N} \boldsymbol{W}_i' \, \boldsymbol{\Omega}_i^{-1} \, (\boldsymbol{y}_i - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}_i)\right] \tag{24}$$

$$\hat{\sigma}_\varepsilon^2 = \left(\sum_{i=1}^{N} n_i\right)^{-1} \sum_{i=1}^{N} \mathrm{tr}\left[(\boldsymbol{\Omega}_i^{-1})\left(\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i' + \boldsymbol{X}_i\boldsymbol{\Sigma}_{\beta|y_i}\boldsymbol{X}_i'\right)\right] \tag{25}$$

$$\begin{aligned}\hat{\boldsymbol{\omega}} = {} & \sigma_\varepsilon^{-2}\left[\sum_{i=1}^{N} \frac{\partial \mathrm{vec}'\,\boldsymbol{\Omega}_i}{\partial\boldsymbol{\omega}} \left(\boldsymbol{\Omega}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1}\right) \frac{\partial \mathrm{vec}\,\boldsymbol{\Omega}_i}{\partial\boldsymbol{\omega}'}\right]^{-1} \\ & \times \sum_{i=1}^{N} \frac{\partial \mathrm{vec}'\,\boldsymbol{\Omega}_i}{\partial\boldsymbol{\omega}} \mathrm{vec}\left[\boldsymbol{\Omega}_i^{-1}\left(\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i' + \boldsymbol{X}_i\boldsymbol{\Sigma}_{\beta|y_i}\boldsymbol{X}_i'\right)\boldsymbol{\Omega}_i^{-1}\right]\end{aligned} \tag{26}$$

where, the "vec" operator stacks the column vectors of a matrix one on top of each other to form one large column vector, "tr" denotes the trace of a matrix, and $\otimes$ denotes the Kronecker product. The equations simplify to some degree in the case of independent errors (*i.e.*, $\boldsymbol{\Omega} = \boldsymbol{I}$ and $\boldsymbol{\omega} = \boldsymbol{0}$).

An EM solution can then be implemented. This solution proceeds by assigning starting values for the structural and population parameters in order to estimate the individual statistics $\tilde{\boldsymbol{\beta}}_i$ and $\boldsymbol{\Sigma}_{\beta|y_i}$ using equations (2) and (3), respectively. Then, these individual statistics are used with the above equations (23)-(26) to obtain improved parameter estimates. This process is repeated until convergence, which can be very slow.

Various authors (Lindstrom and Bates, 1988; Longford, 1993; Bock, 1989b; Hedeker and Gibbons, 1996b) have suggested a Fisher scoring procedure utilizing the first derivatives and expected values of the second derivatives to obtain improved parameter estimates. For this, provisional estimates for the total vector of parameters $\boldsymbol{\Theta}$, on iteration ι are improved by

$$\boldsymbol{\Theta}_{\iota+1} = \boldsymbol{\Theta}_{\iota} - \mathcal{E}\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_{\iota}\, \partial \boldsymbol{\Theta}_{\iota}'}\right]^{-1} \frac{\partial \log L}{\partial \boldsymbol{\Theta}_{\iota}} \,. \tag{27}$$

The negative of the expected values of the second derivatives corresponds to the information matrix (see Bock, 1989b; Hedeker, 1989; and Longford, 1993). At convergence, the large-sample variance covariance matrix of the parameter estimates is then obtained as the inverse of the information matrix.

Although the Fisher scoring solution is a significant improvement in terms of speed of convergence over the EM solution used by Laird and Ware (1982) and others, it can fail in the estimation of the covariance matrix of the random effects as these terms become very small. Lindstrom and Bates (1988) suggest reparameterizing the variance covariance matrix $\boldsymbol{\Sigma}_{\beta}$ in terms of the Cholesky factorization, however, a better choice is to reparameterize in terms of the Gaussian factorization of a symmetric matrix ($\boldsymbol{\Sigma}_{\beta} = \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}'$; see Bock, 1975 pages 82-84), utilizing the exponential transformation for the diagonal matrix $\boldsymbol{D}$ corresponding to the variance parameters. This method is implemented in the MIXREG program (Hedeker & Gibbons, 1996b).

Categorical Response Data. We will present estimation for the ordinal model since the binary model is a special case of the former. For the models in this article, autocorrelated errors have not been considered (*i.e.*, $\boldsymbol{\omega} = \mathbf{0}$) and the error variance has been fixed equal to 1 and $\pi^2/3$ for the probit and logistic formulations, respectively. Otherwise, the mixed-effects regression model for the latent response strength y_{ij},

$$y_{ij} = \boldsymbol{w}_{ij}'\boldsymbol{\alpha} + \boldsymbol{x}_{ij}'\boldsymbol{\beta}_i + \varepsilon_{ij} \tag{28}$$

is the same as for the normal-theory model. Assuming the mixed-effects regression model for the latent response strength, the probability of a response in category k for a given level-2 unit i, conditional on the thresholds (and threshold-varying regression coefficients) $\boldsymbol{\gamma}_k$, regression coefficients $\boldsymbol{\alpha}$, and (standardized) random effects $\boldsymbol{\theta}$ (*i.e.*, $\boldsymbol{\beta} = \boldsymbol{\Lambda}\boldsymbol{\theta}$ where $\boldsymbol{\Sigma}_{\beta} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$), is given by:

$$\Pr(Y_{ij} = k \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda}) = P_{ij\,k} - P_{ij\,k-1} \tag{29}$$

where, for the probit formulation $P_{ij\,k} = \Phi(z_{ijk})$ with $\Phi(\cdot)$ representing the cumulative standard normal density function, and for the logistic response function, $P_{ij\,k} = (1 + \exp(-z_{ijk}))^{-1}$. Here, $z_{ijk} = \boldsymbol{u}_{ij}'\boldsymbol{\gamma}_k + \boldsymbol{x}_{ij}'\boldsymbol{\Lambda}\boldsymbol{\theta}_i + \boldsymbol{w}_{ij}'\boldsymbol{\alpha}$. Also, note that $P_{j\,0} = 0$ and $P_{j\,k} = 1$.

Letting $\mathbf{Y}_{\boldsymbol{i}}$ denote the vector pattern of ordinal item responses from level-2 unit i for the n_i level-1 units nested within, the probability of any pattern $\mathbf{Y}_{\boldsymbol{i}}$, given $\boldsymbol{\gamma}_k$, and $\boldsymbol{\alpha}$, and $\boldsymbol{\theta}$ is equal to the product of the probabilities of the level-1 responses:

$$\ell(\mathbf{Y}_{\boldsymbol{i}} \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda}) = \prod_{j=1}^{n_i} \prod_{k=1}^{K} (P_{ij\,k} - P_{ij\,k-1})^{c_{ijk}} \tag{30}$$

$$\text{where } c_{ijk} = \begin{cases} 1 & \text{if } Y_{ij} = k \\ 0 & \text{if } Y_{ij} \neq k \end{cases} .$$

The marginal density of $\mathbf{Y_i}$ in the population is then expressed as the following integral of the likelihood, $\ell_i = \ell(\mathbf{Y_i} \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda})$, weighted by the prior:

$$h(\mathbf{Y_i}) = \int_{\boldsymbol{\theta}} \ell_i \, g(\boldsymbol{\theta}) \, d\boldsymbol{\theta} \tag{31}$$

where $g(\boldsymbol{\theta})$ represents the multivariate standard normal density.

Just as in the previous normal-theory model, for estimation of the covariate coefficients $\boldsymbol{\alpha}$, the population variance-covariance parameters $\boldsymbol{\Lambda}$, and the threshold parameters $\boldsymbol{\gamma}_k$ $(k = 1, \ldots, K-1)$, the marginal log-likelihood (for the patterns from the N level-2 units) is maximized. Here, this is given as:

$$\log L = \sum_{i}^{N} \log h(\mathbf{Y_i}) \, .$$

Differentiating first with respect to the parameters that vary with k, we get for a particular $\boldsymbol{\gamma}_{k'}$,

$$\frac{\partial \log L}{\partial \boldsymbol{\gamma}_{k'}} = \sum_{i=1}^{N} h^{-1}(\mathbf{Y_i}) \int_{\boldsymbol{\theta}} \sum_{j=1}^{n_i} \sum_{k=1}^{K} c_{ijk} \left[\frac{(\partial P_{ijk}) a_{k\,k'} - (\partial P_{ij\,k-1}) a_{k-1\,k'}}{P_{ijk} - P_{ij\,k-1}} \right] \ell_i \, g(\boldsymbol{\theta}) \, \boldsymbol{u}_{ij} \, d\boldsymbol{\theta} \, , \tag{32}$$

where

$$a_{k\,k'} = \begin{cases} 1 & \text{if } k = k' \\ 0 & \text{if } k \neq k' \end{cases} .$$

Similarly, letting $\boldsymbol{\eta}$ represent an arbitrary parameter vector, then for $\boldsymbol{\alpha}$ and the vector $\mathrm{v}(\boldsymbol{\Lambda})$ which contains the unique elements of the Cholesky factor $\boldsymbol{\Lambda}$, we get:

$$\frac{\partial \log L}{\partial \boldsymbol{\eta}} = \sum_{i=1}^{N} h^{-1}(\mathbf{Y_i}) \int_{\boldsymbol{\theta}} \sum_{j=1}^{n_i} \sum_{k=1}^{K} c_{ijk} \left[\frac{\partial P_{ijk} - \partial P_{ij\,k-1}}{P_{ijk} - P_{ij\,k-1}} \right] \ell_i \, g(\boldsymbol{\theta}) \, \frac{\partial z_{ijk}}{\partial \boldsymbol{\eta}} \, d\boldsymbol{\theta} \, , \tag{33}$$

where

$$\frac{\partial z_{ijk}}{\partial \boldsymbol{\beta}} = \boldsymbol{x}_{ij} \, , \qquad \frac{\partial z_{ijk}}{\partial (\mathrm{v}(\boldsymbol{\Lambda}))} = \mathsf{J}_r(\boldsymbol{\theta} \otimes \boldsymbol{w}_{ij}) \, ,$$

and J_r is the transformation matrix of Magnus (1988) that eliminates the elements above the main diagonal. Note that for the logistic formulation, $\partial P_{ijk} = P_{ijk}(1 -$

P_{ijk}), while for the probit it is obtained using the standard normal density function $\phi(\cdot)$.

As in the normal-theory model, Fisher's method of scoring can be used to provide the solution to these likelihood equations. Here, the empirical information matrix can be used:

$$\mathcal{E}\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_\iota \, \partial \boldsymbol{\Theta}_\iota'}\right] = -\sum_{i=1}^{N} h^{-2}(\mathbf{Y}_i) \, \frac{\partial h(\mathbf{Y}_i)}{\partial \boldsymbol{\Theta}_\iota} \left(\frac{\partial h(\mathbf{Y}_i)}{\partial \boldsymbol{\Theta}_\iota}\right)' . \tag{34}$$

The derivation of the empirical information matrix follows directly from Bock and Lieberman (1970). As noted by Bock and Aitkin (1981), this empirical information matrix will in general be positive-definite provided that the number of level-2 units N exceeds the number of parameters to be estimated.

Numerical Quadrature. In order to solve the above likelihood equations, numerical integration on the transformed $\boldsymbol{\theta}$ space can be performed. For this, Gauss-Hermite quadrature can be used to approximate the above integrals to any practical degree of accuracy. In Gauss-Hermite quadrature, the integration is approximated by a summation on a specified number of quadrature points Q for each dimension of the integration; thus, for the transformed $\boldsymbol{\theta}$ space, the summation goes over Q^r points. For the standard normal univariate density, optimal points and weights (which will be denoted B_q and $A(B_q)$, respectively) are given in Stroud and Sechrest (1966). For the multivariate density, the r-dimensional vector of quadrature points is denoted by $\boldsymbol{B_q}' = (B_{q1}, B_{q2}, \ldots, B_{qr})$, with its associated (scalar) weight given by the product of the corresponding univariate weights,

$$A(\boldsymbol{B_q}) = \prod_{h=1}^{r} A(B_{qh}) \, . \tag{35}$$

If another distribution is assumed, other points may be chosen and density weights substituted for $A(B_q)$ or $A(B_{qh})$ above. For each of the r dimensions, these weights must be normalized to sum to unity. For example, if a rectangular or uniform distribution is assumed, then Q points may be set at equal intervals over an appropriate range (for each dimension) and the quadrature weights are then set equal to $1/Q$. Other distributions are possible: Bock and Aitkin (1981) discuss the possibility of empirically estimating the random-effect distribution.

For models with few random effects the quadrature solution is relatively fast and computationally tractable. In particular, if there is only one random effect in the model, there is only one additional summation over Q points relative to the fixed effects solution. As the number of random effects r is increased, the terms in the summation (Q^r) increases exponentially in the quadrature solution. Fortunately, as is noted by Bock, Gibbons and Muraki (1988) in the context of a binary factor analysis model, the number of points in each dimension can be reduced as the dimensionality is increased without impairing the accuracy of the approximations; they indicated that for a five-dimensional solution as few as three points per dimension were sufficient to obtain adequate accuracy.

At each iteration and for each level-2 unit, the solution goes over the Q^r quadrature points, with summation replacing the integration over the random-effect distribution. The conditional probabilities $\ell(\mathbf{Y_i} \mid \boldsymbol{\theta}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda})$ are obtained substituting the random-effect vector $\boldsymbol{\theta}$ by the current r-dimensional vector of quadrature points $\boldsymbol{B_q}$. The marginal density for each level-2 unit is then approximated as

$$h(\mathbf{Y_i}) \approx \sum_{q}^{Q^r} \ell(\mathbf{Y_i} \mid \boldsymbol{B_q}; \boldsymbol{\gamma}_k, \boldsymbol{\alpha}, \boldsymbol{\Lambda})\ A(\boldsymbol{B_q})\ . \tag{36}$$

At each iteration, computation of the first derivatives and information matrix then proceeds summing over level-2 units and quadrature points. In the summation over the Q^r quadrature points, substitutions are made in the equations for the first derivatives and information matrix as follows: the $\boldsymbol{\theta}$ random-effect vector is substituted by the current vector of quadrature points $\boldsymbol{B_q}$, and the evaluation of the multivariate standard normal density $g(\boldsymbol{\theta})$ is substituted by the current quadrature weight $A(\boldsymbol{B_q})$. Following the summation over level-2 units and quadrature points, parameters are corrected using (27), and the entire procedure is repeated until convergence. The method described here is implemented in the MIXOR program (Hedeker and Gibbons, 1996a).

References

Agresti, A., and Lang, J.B. (1993). A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika*, **80**, 527-534.

Albert, J.H., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Jour. Amer. Statist. Assoc.*, **88**, 669-680.

Ashford, J.R., and Sowden, R.R. (1970). Multivariate probit analysis. *Biometrics,* **26**, 535-546.

Bliss, C.J. (1935). The calculation of the dosage mortality curve. *Ann. Appl. Biol.*, **22**, 307-330.

Bock, R.D., and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika,* **35**, 179-197.

Bock, R.D. (1975). *Multivariate Statistical Methods in Behavioral Research*, New York, McGraw-Hill.

– – –– (1979). Univariate and multivariate analysis of time-structured data. In: Nesselroade JR, Baltes PB, eds. *Longitudinal Research in the Study of Behavior and Development.* New York, NY: Academic Press.

Bock, R.D., and Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika,* **46**, 443-459.

Bock, R.D. (1983). The discrete Bayesian. In H. Wainer & S. Messick (eds.), *Principles of Modern Psychological Measurement.* Hillsdale, NJ: Earlbaum.

Bock, R.D., Gibbons, R.D., and Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement,* **12**, 261-280.

Bock, R. D. (Ed.). (1989a) *Multilevel Analysis of Educational Data*, New York: Academic Press.

– – –– (1989b). Measurement of human variation: a two stage model. In R.D. Bock (ed.), *Multilevel Analysis of Educational Data.* New York: Academic Press.

Bock, R.D., and Gibbons R.D. (1996). High dimensional multivariate probit analysis. *Biometrics*, **52**, 1183-1194.

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

Bryk, A.S., and Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*, Newbury Park, CA: Sage Publications, Inc..

Bryk, A.S., Raudenbush, S.W., and Congdon, R.T. (1996). *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs.* Chicago: Scientific Software International.

Chi, E.M., and Reinsel, G.C. (1989). Models of longitudinal data with random effects and AR(1) errors. *Jour. Amer. Statist. Assoc.*, **84**, 452-459.

Cox, C. (1995). Location-scale cumulative odds models for ordinal data: a generalized non-linear model approach. *Statistics in Medicine*, **14,** 1191-1203.

DeLeeuw, J., and Kreft, I. (1986). Random coefficient models for multilevel analysis. *Jour. Educational Statistics*, **11**, 57-85.

Dempster, A.P., Rubin, D.B., and Tsutakawa, R.K. (1981). Estimation in covariance component models. *Jour. Amer. Statist. Soc.*, **76**, 341-353.

Diggle, P., and Kenward M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49-94.

Diggle, P., Liang, K.-Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, New York: Oxford University Press.

Egret (1991). *Reference Manual.* Seattle, WA: Statistics and Epidemiology Research Corporation.

Elston, R.C., and Grizzle, J.E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics*, **18**, 148-159.

Ezzet, F., and Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine*, **10**, 901-907.

Fay, J.W.J. (1957). The National Coal Board's pneumoconiosis research. *Nature*(London), **180**, 309.

Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika*, **62**, 89-100.

Fitzmaurice, G.M., Laird, N.M, and Rotnitzky, A.G. (1993). Regression models for discrete longitudinal responses. *Statistical Science*, **8**, 284-309.

Gibbons, R.D., and Bock, R.D. (1987). Trend in correlated proportions. *Psychometrika*, **52**, 113-124.

Gibbons R.D., Hedeker D., Elkin I., Waternaux C., Kraemer H.C., Greenhouse J.B., Shea M.T., Imber S.D., Sotsky S.M., and Watkins J.T. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry*, **50**, 739-750.

Gibbons, R.D., and Hedeker, D. (1994). Application of Random effects probit regression models. *Jour. Clinical and Consulting Psychology*, **62**, 285-296.

Gibbons, R.D., Hedeker, D., Charles, S.C., and Frisch P. (1994). A random-effects probit model for predicting medical malpractice claims. *Jour. Amer. Statist. Assoc.*, **89**, 760-767.

Gibbons R.D., and Wilcox-Gök V. (1998). Health service utilization and insurance coverage: A multivariate probit analysis. *Jour. Amer. Statist. Assoc.*, **93**, 63-72.

Gibbons R.D., and Hedeker D. (1997). Random-effects probit and logistic regression models for three-level data. *Biometrics*, **53**, 1527-1537.

Gibbons R.D., and Lavigne J.V. (1998). Emergence of childhood psychiatric disorders: A multivariate probit analysis. *Statistics in Medicine*, **17**, 2487-2499.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, **73**, 43-56.

– – ––– (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, **78**, 45-51.

– – – – (1995). *Multilevel Statistical Models* (2nd edition). New York: Halstead Press.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *Jour. Amer. Statist. Assoc.,* **72,** 320-385.

Harville, D.A., and Mee, R.W. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics,* **40**, 393-408.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Ann. Economic and Social Measurement*, **5**, 475-492.

HEDEKER, D. (1989). Random regression models with autocorrelated errors. *Ph.D. thesis*, University of Chicago.

HEDEKER, D., AND GIBBONS, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, **50**, 933-944.

HEDEKER D., AND GIBBONS R.D. (1996a). MIXOR: a computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, **49**, 157-176.

HEDEKER D., AND GIBBONS R.D. (1996b). MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine*, **49**, 229-252.

HEDEKER D., AND GIBBONS R.D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, **2**, 64-78.

HEDEKER D., AND MERMELSTEIN, R.J. (1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, **33**, 427-455.

JAMES, W., AND STEIN, C. (1961). Estimation with quadratic loss. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 361-379.

JANSEN, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics,* **39**, 75-84.

JOHNSON, V.E. (1996). On Bayesian analysis of multirater ordinal data: An application to automated essay grading. *Jour. Amer. Statist. Assoc.*, **91**, 42-51.

JONES, R.H. (1993). *Longitudinal Data Analysis with Serial Correlation: a State-space Approach*, New York: Chapman and Hall.

KOCH, G., LANDIS, J., FREEMAN, J., FREEMAN, H., AND LEHNEN, R. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, **33**, 133-158.

KREFT, I.G., DE LEEUW, J., AND VAN DER LEEDEN, R. (1994). Comparing five different statistical packages for hierarchical linear regression: BMDP-5V, GENMOD, HLM, ML3, and VARCL. *American Statistician*, **48**, 324-335.

LAIRD, N.M., AND WARE, J.H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.

LAIRD, N.M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, **7**, 305-315.

LIANG, K.Y., AND ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

LINDLEY, D.V., AND SMITH, A.F.M. (1972). Bayes estimation for linear models(with Discussion). *J. Roy. Statist. Soc.*,Series B, **34**, 1-41.

LINDSEY, J.K. (1993). *Models for Repeated Measurements.* New York, Oxford University Press.

LINDSTROM, M.J., AND BATES, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Jour. Amer. Statist. Assoc.*, **404**, 1014-1022.

LITTLE, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Jour. Amer. Statist. Soc.*, **88**, 125-133.

– – – – (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471-483.

– – – – (1995). Modeling the drop-out mechanism in repeated-measures studies. *Jour. Amer. Statist. Assoc.*, **90**, 1112-1121.

LITTLE, R.J.A., AND RUBIN D.B. (1987). *Statistical Analysis with Missing Data* New York, Wiley.

LONGFORD, N.T. (1986). VARCL - Interactive software for variance component analysis: Applications for survey data. *Professional Statistician*, **5**, 28-32.

– – – – (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects, *Biometrika*, **74**, 817-827.

– – – – (1993). *Random Coefficient Models.* New York, Oxford University Press.

LORD, F.M., AND NOVICK, M.R. (1968). *Statistical Theories of Mental Test Scores.* Reading, Mass: Addison-Wesley.

MAGNUS, J. R. (1988). *Linear structures.* London: Charles Griffin.

MANSOUR, H. NORDHEIM E.V., AND RUTLEDGE J.J. (1985). Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors, *Biometrics*, **41**, 287-294.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc.*, Series B, **42**, 109-142.

Muthén, B. (1979). A structural probit model with latent variables. *Jour. Amer. Statist. Assoc.*, **74**, 807-811.

Neuhaus, J.M., Kalbfleisch, J.D., and Hauck, W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59**, 25-35.

Normand, S.L. Glickman, M.E., and Ryan, T. (1997). Modeling mortality rates for elderly heart attack patients: Profiling hospitals in the Cooperative Cardiovascular Project. In Gastonis, C., Hodges, J., Kass, R. & Singpurwalla, N. (eds.), em Case Studies in Bayesian Statistics, Springer-Verlag.

Pendergast, J.F., Gange, S.J, Newton, M.A., Lindstrom, M.J., Palta, M., and Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review*, **64**, 89-118.

Peterson, B., and Harrell, F.E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, **39**, 205-217.

Rasbash, J., Wang, M., Woodhouse, G., and Goldstein, H. (1995). *MLn: Command Reference Guide*. London: Institute of Education, University of London.

Rosenberg, B. (1973). Linear regression with randomly dispersed parameters. *Biometrika*, **60**, 65-72.

Schluchter, M.D. (1988). Unbalanced repeated measures models with structured covariance matrices. In: W.J. Dixon (Chief Ed.), *BMDP Statistical Software Manual* (vol. 2) (pp. 1081-1114). Berkeley, CA: University of California Press.

Searle, S.R. Casella, G., and McCulloach, C.E. (1992). *Variance Components*, New York: Wiley.

Seltzer, M.H., Wong, W.H., and Bryk, A.S. (1996). Bayesian analysis in applications of hierarchical models: issues and methods. *Journal of Educational and Behavioral Statistics*, **21**, 131-167.

Stiratelli, R., Laird, N.M., and Ware, J.H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, **40**, 961-971.

Stroud, A.H., and Sechrest, D. (1966). *Gaussian Quadrature Formulas.* Englewood Cliffs, NJ: Prentice Hall.

Ten Have, T.R. (1996). A mixed effects model for multivariate ordinal response data including correlated failure times with ordinal responses. *Biometrics*, **52,** 473-491.

Terza, J.V. (1985). Ordinal probit: a generalization. *Communications in Statistical Theory and Methods*, **14,** 1-11.

Wong, G.Y., and Mason, W.M. (1985). The hierarchical logistic regression model for multilevel analysis. *Jour. Amer. Statist. Assoc.*, **80**, 513-524.

Zeger, S.L., and Liang, K-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics,* **42**, 121-130.

Robert D. Gibbons
Department of Psychiatry and
Division of Epidemiology & Biostatistics
University of Illinois at Chicago
Biometric Laboratory
1601 West Taylor Street
Chicago, IL 60612, U.S.A.
E-mail: robert.gibbons@uic.edu

Donald Hedeker
Division of Epidemiology & Biostatistics
and Health Research & Policy Centers
University of Illinois at Chicago
2121 West Taylor Street, Room 525
Chicago, IL 60612-7260, U.S.A.
E-mail: hedeker@uic.edu