Sankhyā : The Indian Journal of Statistics
Special Issue on Biostatistics
2000, Volume 62, Series B, Pt. 1, pp. 134–148

LONGITUDINAL DATA ANALYSIS OF CONTINUOUS AND DISCRETE RESPONSES FOR PRE-POST DESIGNS*

By KUNG–YEE LIANG
and
SCOTT L. ZEGER
Johns Hopkins University

SUMMARY. The availability of biomarkers has led to the increasing adoption of the pre- and post-randomization designs in clinical trials. In this paper we discuss the use of random effects models when the primary objective of the trial is to assess the efficacy of newly developed treatments in terms of progression of biomarkers over time. One issue of particular interest is how to best utilize the pre-randomization responses. We discuss the pros and cons of several inferential procedures including the conditional and full likelihood approaches. Throughout, these issues are illustrated by an analysis of data from a schizophrenic trial by the Janssen Research Foundation.

1. Introduction

Conventional clinical trials compare time to a pre-specified event for patients randomly assigned to different treatments. Such an event could be death for terminally ill cancer patients or the time to hospitalization for schizophrenia among subjects. Recent rapid development of molecular medicine has led to the use of biological markers as the outcome variable for comparison. Examples include CD4 cell counts, serum cholesterol levels or psychiatric symptom scores. In addition, the concept of quality of life (QOL) has recently gained currency as an additional response to death when the targeted population is terminally ill patients. The availability and use of these "markers", which can be measured repeatedly, has led to the increasing adoption of the so-called "pre- and post-randomization designs" in clinical trials.

This paper discusses the use of longitudinal data analysis methods (Diggle, Liang and Zeger, 1994) for the analysis of data from such trials. Consider a recent study by the Janssen Research Foundation on the effectiveness of risperidone for subjects suffering from schizophrenia (Chouinard *et al.*, 1993; Marder and Meiback, 1994).

AMS (1991) *subject classification.* Primary 62J10, 62J12.
Key words and phrases. Clinical trials; longitudinal data; pre-post designs; random effects model
* The authors gratefully acknowledge support from NIH grants MH 56639 and CM 49909 as well as partial support from Janssen Pharmaceutica.

Six hundred patients were randomized into six arms: placebo, the standard treatment known as haloperidol (20 mg) and a new treatment, risperidone, at four doses: 2, 6, 10 and 16 mg. The primary response variable is the Positive and Negative Score Scale (PANSS), which is a standard measurement of the severity of schizophrenia; the higher the value of PANSS, the more severe the condition. For each participant, the PANSS was measured once prior to the randomization and at 1, 2, 4, 6 and 8 weeks after the randomization. The primary objective of the trial was to determine the optimal dose of risperidone to slow the increase of the PANSS over time.

To establish notation, let Y_{ij} be the response variable for the i-th subject measured at the j-th occasion, $j = 0, 1, \ldots, T$ and $i = 1, \ldots, n$. Here $j = 0$ indexes the pre-randomization measurement and T is the number of post-randomization observations for each subject. Furthermore, let z_i be the dummy variables for treatment status and let x_{ij} be a $p \times 1$ vector of covariates which could be time-independent, perhaps measured only at baseline, or time-dependent including time from baseline.

In this paper, we focus on estimating the changes in Y_{ij} over time and how these changes differ from subjects assigned to different treatment groups. One issue of particular interest is how to best utilize the the pre-randomization response. We discuss the pros and cons of several approaches to inference including the conditional and full likelihood methods.

2. Statistical Models for Continuous Responses

With $T + 1$ observations from each subject, one can model his or her change in the response variable over time using a simple linear regression

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_j + \epsilon_{ij}, \quad j = 0, 1, \ldots T. \tag{1}$$

Here t_j is the time interval between the j-th visit and the baseline. Note by definition $t_o \equiv 0$; thus β_{0i} represents the expected baseline measurement of the response variable for person i, i.e. $\beta_{0i} = \mathrm{E}(Y_{i0})$ and β_{1i} is the expected rate of change in Y_{ij}. Note that the model in (1) may be easily extended if a linear relationship is not appropriate by adding higher order terms in t_j, e.g. t_j^2 or replacing t_j by a more appropriate function in t_j, e.g. $\log(t_j + 1)$. Finally, $\{\epsilon_{ij}, j = 0, \ldots T\}$, the error terms, are assumed to be independent of each other and normally distributed with common mean zero and variance σ^2.

Recall that the primary objective of the trial is to examine how individual change in Y varies among differential treatment groups. This may be achieved by modeling β_{1i} as

$$\beta_{1i} = \gamma_0 + \gamma_1' z_i + b_{1i}, \quad i = 1, \ldots, n, \tag{2}$$

where z_i is a vector of treatment group indicator variables, $z_i = 1$ if person i is in treatment group, and 0 otherwise. Then, γ_1 is the difference in average rate of change between each treatment group and the control group, and b_{1i} the deviation of each subject's change rate from his group's average.

As in (2), one may also model β_{0i} of (1) through regression, i.e.

$$\beta_{0i} = \alpha_0 + \alpha_1' z_i + \alpha_2' x_i + b_{oi},\ i = 1, \ldots, n. \tag{3}$$

Here z_i and x_i are included in (3) to allow for the possible difference in $\mathrm{E}(Y_{i0})$ among treatment groups and to adjust for baseline covariates x_i. Note that z_i in (2) is the main covariate of interest. However, it is not necessary that z_i be included in (3). This is because in a well conducted randomized trial, one would expect that, on average, the Y_{i0}s would be comparable among treatment groups. This issue of whether to adjust for z_i in Y_{i0} is an example of the typical trade-off between bias and precision in statistical inference. This will be discussed in greater detail in §3.

To complete the modeling process, it is typical, although not entirely necessary, to assume that $(b_{0i}, b_{1i})'$ are independent of $\{\epsilon_{ij}, j = 0, \ldots, T\}$ and are generated from a common distribution, e.g.

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \mathrm{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \right). \tag{4}$$

By substituting (2) and (3) for β_{1i} and β_{0i}, respectively, in (1), we have

$$Y_{ij} = \alpha_0 + \alpha_1' z_i + \alpha_2' x_i + \gamma_0 t_j + \gamma_1' z_i t_j + \epsilon_{ij}^*, \tag{5}$$

where $\epsilon_{ij}^* = \epsilon_{ij} + b_{0i} + b_{1i} t_j$. Thus, parameters of interest, γ_1 appear in (5) as the regression coefficients for the interactions between z_i and t_j.

Statistical models (1)-(4) for pre- and post-randomization designs are known to be a special case of the random effects models and have been applied to educational testing and social science research (e.g. Allison, 1990) and controlled clinical trials (e.g. Laird and Wang, 1990). One important feature of this approach is that Y_{i0} is treated like $\{Y_{i1}, \ldots, Y_{iT}\}$ as a response variable rather than being considered as a covariate. Another aspect, as will be seen in detail in §4, is that this approach can be extended easily to models for non-Gaussian observations. We also note that assumptions (1)-(4), fully specify the data likelihood so that maximum likelihood estimation is readily available for $\theta_1 = (\gamma_0, \gamma_1), \theta_2 = (\alpha_0, \alpha_1', \alpha_2')$ and $\theta_3 = (\sigma^2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22})$. We will also discuss in §3 and 4 an alternative approach using conditional likelihoods.

3. Estimating Procedures for Continuous Responses

We first consider the simple case in which $T = 1$, i.e. there is only one post-randomization observation ($T = 1$) and two treatment groups. Thus $z_i = 1(0)$ if the i-th subject is randomized to the new treatment (placebo). In addition, we assume, without loss of generality, that $t_1 = 1$. In this special case, a closed form expression for estimators of γ_1 are available which helps to illuminate some of the issues we address later.

3.1 *One post-randomization response.* When there are only two observations (Y_{i0}, Y_{i1}) per subject, one encounters the issue of over-parameterization for the covariance parameters $\theta_3 = (\sigma^2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22})$. One simple solution is to assume $\Sigma_{22} = \mathrm{Var}(b_{1i}) = 0$ and hence the b_{1i} are degenerate at zero. Under this assumption, the likelihood function for $\theta = (\theta_1, \theta_2, \theta_3)$ based on $Y_i = (Y_{i0}, Y_{i1})$ may be expressed as

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n} \int f(y_i \mid b_{0i}; \theta_1, \theta_2, \sigma^2) f(b_{0i}; \Sigma_{11}) db_{0i} \\ &= \prod_{i=1}^{n} \int f(y_i \mid y_{i+} = y_{i0} + y_{i1}; \theta_1, \sigma^2) f(y_{i+} \mid b_{0i}; \theta_1, \theta_2, \sigma^2) f(b_{0i}; \Sigma_{11}) db_{0i} \\ &= \prod_{i=1}^{n} f(y_i \mid y_{i+}; \theta_1, \sigma^2) f(y_{i+}; \theta), \end{aligned} \tag{6}$$

where f is an exponential family probability density function. As indicated in (6), the conditional distribution of Y_i given y_{i+} is independent of both b_{0i} and θ_2. Independence of b_{0i} is a simple consequence of the fact that $f(y_i \mid b_{0i})$ is from an exponential family (Gaussian in this case) with $Y_{i+} = Y_{i0} + Y_{i1}$ being the sufficient statistic for b_{0i}. To see the independence of θ_2, note that $\theta_2 = (\alpha_0, \alpha_1', \alpha_2')$ are coefficients for covariates which are subject-specific (z_i and x_i) and hence the sufficient statistic for θ_2 based on Y_i is totally confounded with that for b_{0i}, Y_{i+}. More specifically, the conditional likelihood for θ_1 and σ^2 based on $f(y_i \mid y_{i+})$ has the form

$$L_C(\theta_1, \sigma^2) \propto \prod_{i=1}^{n} f(y_i \mid y_{i+}; \theta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{4\pi\sigma^2}} \mathrm{e}^{(d_i - \gamma_0 - \gamma_1' z_i)^2 / 4\sigma^2} \tag{7}$$

where $d_i = y_{i1} - y_{i0}$, the change in Y for the i-th subject. It can be seen easily that the estimator which maximizes the conditional likelihood L_C is simply the difference in the averaged ds for $z \equiv 1$ versus $z \equiv 0$, i.e.

$$\hat{\gamma}_{1C} = \frac{\sum_{i=1}^{n} z_i d_i}{n_1} - \frac{\sum_{i=1}^{n} (1 - z_i) d_i}{n_0} = \bar{d}^{(1)} - \bar{d}^{(0)},$$

where $n_1 = \sum_{i=1}^{n} z_i = n - n_0$.

This conditional likelihood approach provides a justification for the intuitively appealing approach of using d_i as the basis for inference about γ_0 and γ_1. This approach is robust in that it is invariant to (i) the distributional assumption on the b_{0i}s which are treated as fixed parameters, and (ii) how β_{0i} in (3) is modeled. In

contrast, the full likelihood approach requires the specification of $f(b_{0i})$ as seen in (6). More importantly, the form for the maximum likelihood estimator (MLE) of $\gamma_1, \hat{\gamma}_1$, depends critically on how β_{0i} is modeled. For example, in the absence of baseline covariate effects, i.e. $\alpha_2 \equiv 0, \hat{\gamma}_1$ is equal to $\hat{\gamma}_{1C}$ if z_i is included in (3) and is equal to

$$\tilde{\gamma}_1 = \hat{\gamma}_{1C} + \frac{\sigma^2}{\Sigma_{11} + \sigma^2}(\bar{y}_0^{(1)} - \bar{y}_0^{(0)}),$$

if z_i is omitted in (3), where $\bar{y}_0^{(1)}$ and $\bar{y}_0^{(0)}$ are the sample means in baseline for $z \equiv 1$ and 0, respectively, (Stanek, 1988). It is easy to see (e.g. Stanek, 1988) that under (2) and (3)

$$\mathrm{E}(\tilde{\gamma}_1) = \gamma_1 + \frac{\sigma^2}{\Sigma_{11} + \sigma^2}\alpha_1. \tag{8}$$

Thus, except when $\alpha_1 = 0$, $\tilde{\gamma}_1$ is a biased estimator of γ_1, whereas $\hat{\gamma}_{1C}$ is unbiased whether $\alpha_1 = 0$ or not. On the other hand, because of regression to the mean when $\alpha_1 = 0$, $\tilde{\gamma}_1$ is more precise than $\hat{\gamma}_{1C}$ as

$$\mathrm{Var}(\tilde{\gamma}_1) = \left(\frac{2n\sigma^2}{n_1 n_0}\right)\left(1 - \frac{\sigma^2}{2(\Sigma_{11} + \sigma^2)}\right) = \mathrm{Var}(\hat{\gamma}_{1C})\left(1 - \frac{\sigma^2}{2(\Sigma_{11} + \sigma^2)}\right).$$

The ratio of these two variance may be re-expressed as $(1 + \rho)/2$ where $\rho = \Sigma_{11}/(\Sigma_{11} + \sigma^2)$ is the within-subject correlation. For many biomarkers, ρ is relatively high ($\geq .7$, say). Thus the gain in efficiency by using $\tilde{\gamma}_1$ appears to be small in many instances. In light of this bias and precision trade-off between the two approaches, we recommend that both approaches be taken for analysis. Qualitatively, similar results from both approaches helps to strengthen the common conclusions drawn from the analysis.

These two estimators, $\hat{\gamma}_{1C}$ and $\tilde{\gamma}_1$, may also be contrasted in a second way. Under the assumptions (1)-(4) and conditional on y_{i0}, $d_i = y_{i1} - y_{i0}$ is normally distributed with mean

$$\gamma_0 + (1 - \rho)\alpha_0 + \{\gamma_1 + (1 - \rho)\alpha_1)\}z_i + (\rho - 1)y_{i0} = \gamma_0^* + \gamma_1^* z_i + \gamma_2^* y_{i0} \tag{9}$$

and variance $\sigma^2 \times (2 - (1 - \rho)^2)$. With straightforward algebra, one can show that the MLE for γ_1^* in (9) is equal to $\tilde{\gamma}_1$. The relationship between γ_1^* and γ_1 explains why $\tilde{\gamma}_1$ is biased as expressed in (8) except when $\alpha_1 = 0$. Noting the relationship between γ_1^* and γ_2^*, an unbiased estimator of γ_1 is achieved by regressing on d_i and z_i without adjusting for y_{i0} (i.e. forcing γ_2^* to be zero). This is exactly how $\hat{\gamma}_{1C}$ was derived as shown in (7). It is worth noting that the former approach of simultaneously regressing d_i on z_i and y_{i0} is known as the regressor variable method, whereas the latter approach, regressing d_i on z_i only, is known as the change score method in the social science literature (e.g. Allison, 1990). Collectively, we call this approach of regressing d_i on z_i and/or y_{i0} the ANCOVA approach for simplicity.

For illustration, we return to the schizophrenic trial with the focus on two treatments: $z = 1$ if receiving risperidone of 6mg ($n_1 = 85$) and $z = 0$ if receiving haloperidol of 20 mg ($n_0 = 85$). Figure 1 shows the box plots of PANSS at baseline

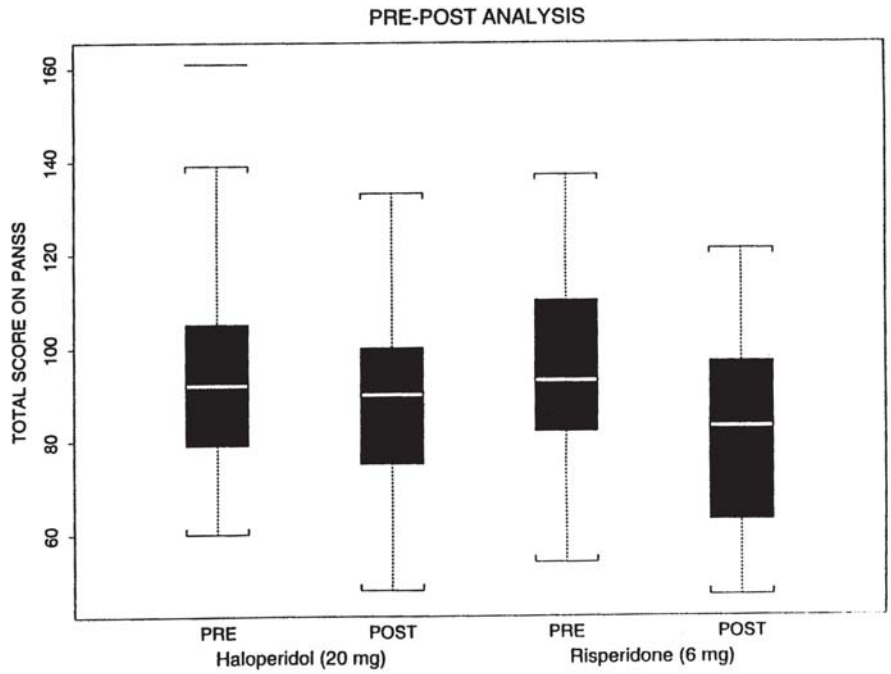

Figure 1: Box plots for pre and post- PANSS scores for Haloperidol (20mg) and Risperidone (6mg) patients.

($t = 0$) and one week after randomization ($t = 1$) for both groups. We note that the treated group ($z = 1$) has a slightly higher PANSS at baseline than the control group ($z = 0$) so that $\hat{\alpha}_1 = 1.26$ with estimated s.e. $= 2.92$. Figure 2 shows the change in PANSS for each subject stratified by treatment status. Table 1 gives estimates of γ_1 for three different approaches: full likelihood, conditional likelihood and two ANCOVA models: Model 1: z_i included in (3); and Model 2: z_i omitted from (3). As expected, the estimates of γ_1 are different with $\hat{\gamma}_1 = 8.77 \pm 2.47$ in Model 1 and $\hat{\gamma}_1 = 8.33 \pm 2.25$ in Model 2 when the full likelihood method is employed. The gain in efficiency by assuming the baseline PANSS are comparable between the two groups, i.e. Model 2, appears to be small with the ratio of $\text{var}(\hat{\gamma}_1)$ being $(2.25/2.47)^2 = 0.83$ as the within-subject correlation ρ is estimated as 0.65. As previously shown, the ANCOVA method nearly gives identical estimates to the full likelihood approach. The slight discrepancy is due to different approaches for estimating σ^2: the full likelihood method uses both Y_{i0} and Y_{i1} to estimate σ^2, whereas the ANCOVA method only uses d_i. Finally, results from the conditional likelihood method are the same whether z_i is included or not and are identical to that from the other two approaches when fitting Model 1 to the data. Regardless of the method, the data indicate that there is an eight unit greater decline in PANSS in one week ($\hat{\gamma}_1 \approx -8$) for subjects receiving a risperidone dose of 6mg than subjects receiving haloperidol at 20 mg.

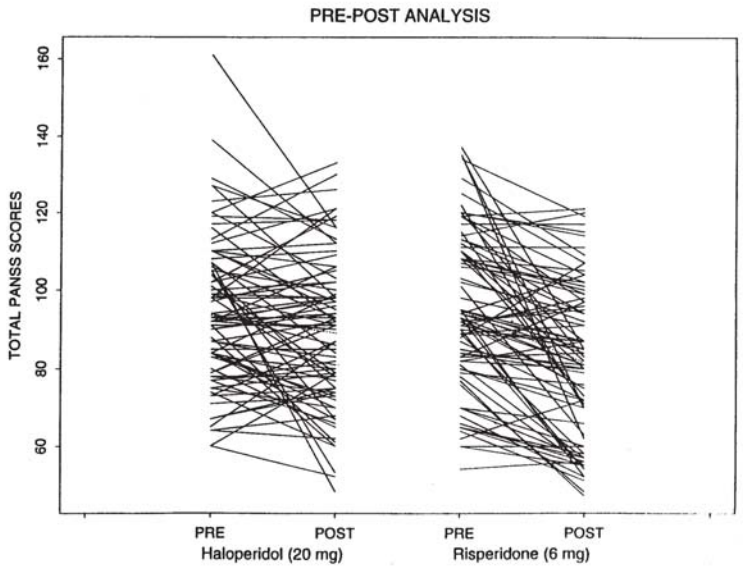

Figure 2: PANSS scores for each patient connected with line segments.

Table 1. Inference for γ_1 based on one post-randomization response ($t_1 = 1$ week) for the schizophrenic trial

Approach	Model 1*	Model 2*
Full likelihood	-8.77	-8.33
	(2.47)	(2.25)
Conditional likelihood	-8.77	-8.77
	(2.47)	(2.47)
ANCOVA		
	Without y_{i0}	with y_{i0}
	-8.78	-8.34
	(2.46)	(2.24)

* Model 1: baseline PANSS is assumed to be different between two groups
Model 2: baseline PANSS is assumed to be the same between two groups

In this section where the focus is on continuous responses with two observations per subject, we have drawn the connection among three approaches: the conditional and full likelihood approach and the ANCOVA approach with d_i as the dependent variable. The pros and cons of the two competing estimators, $\hat{\gamma}_{1C}$ and $\tilde{\gamma}_1$, have been discussed. As will become clear below, the full likelihood approach is our method of choice as the other two approaches are not easily extended to situations with discrete responses and/or random slopes.

3.2 *Multiple post-randomization responses.* The conditional likelihood approach can also be applied to models for multiple post-randomization observations if we assume $\mathrm{Var}(b_{1i}) = 0$. In this case, $Y_{i+} = \sum_{j=0}^{T} Y_{ij}$ remains the sufficient statistic for b_{0i} and the conditional distribution of Y_i given Y_{i+} is

$$f(Y_i = y_i \mid Y_{i+} = y_{i+}) = \left(2\pi\sigma^2 / \sum_{j=1}^{T}(t_j - \bar{t})^2\right)^{-1/2} \exp\left(\frac{-(\hat{\beta}_{1i} - \gamma_0 - \gamma_1' z_i)^2)}{2\sigma^2 / \sum_j (t_j - \bar{t})^2}\right), \tag{10}$$

where $\hat{\beta}_{1i} = \sum_j (t_j - \bar{t})(y_{ij} - \bar{y}_i) / \sum_j (t_j - \bar{t})^2$, $\bar{t} = \sum_j t_j / T$ and $\bar{y}_i = y_{i+}/T$. This leads to the conditional MLE for γ_1

$$\hat{\gamma}_{1C} = \sum_{i=1}^{n} z_i \hat{\beta}_{1i} / n_1 - \sum_{i=1}^{n} (1 - z_i)\hat{\beta}_{1i} / n_0.$$

However, this approach may be criticized on the ground that the assumption that $\mathrm{var}(b_{1i}) = 0$ is unrealistic due to natural heterogeneity in time trends among subjects. In addition, with $T > 1$, the earlier concern of over-parameterization becomes less important. Unfortunately, this conditional likelihood approach, which was motivated to eliminate large numbers of nuisance parameters, cannot be applied in the situation where both b_{0i} and b_{1i} are included. This is because the joint sufficient statistics for (b_{0i}, b_{1i}) are now $(Y_{i+}, \hat{\beta}_{1i})$ which is fully informative for (θ_1, θ_2) as well. Thus, by conditioning on $(Y_{i+}, \hat{\beta}_{1i})$, the distribution of Y_i contains no information on either θ_1 or θ_2, the latter being the parameters of interest.

The full likelihood approach alleviates this problem by imposing the additional assumption (4) that (b_{0i}, b_{1i}) is generated from a common mean-zero distribution indexed by θ_3 so that the likelihood, which is a function of $\theta = (\theta_1, \theta_2, \theta_2)$, can be written as

$$L(\theta) \propto \prod_{i=1}^{n} \int f(y_i \mid b_{0i}, b_{1i}; \theta_1, \theta_2, \sigma^2) f(b_{0i}, b_{1i}; \theta_3) db_{0i} db_{1i}.$$

Straightforward likelihood inference follows especially if the bivariate normality assumption is made for (b_{0i}, b_{1i}) in which case the marginal distribution for Y_i is also multivariately normal with

$$\mathrm{E}(Y_{ij}) = \alpha_0 + \alpha_1' z_i + \alpha_2' x_i + \gamma_0 t_j + \gamma_1' z_i t_j$$

and

$$\begin{aligned} \mathrm{cov}(Y_{ij}, Y_{ik}) &= \Sigma_{11} + t_j^2 \Sigma_{22} + 2 t_j \Sigma_{12} \qquad j < k \\ &= \sigma^2 + \Sigma_{11} + t_j^2 \Sigma_{22} + 2 t_j \Sigma_{12}, \quad j = k. \end{aligned}$$

However, as mentioned earlier, this approach requires the additional specification on the distribution for (b_{0i}, b_{1i}) and the MLE for parameters of interest, γ_1, depends on whether z_i is included in the regression model for the mean (3).

The third approach extends the ANCOVA approach by using d_i as the dependent variable. Here, we regress $\hat{\beta}_{1i}$ on z_i and/or y_{i0}. While intuitive, one should use this approach with caution when partial observations from some of the subjects are missing. First, the constant variance assumption for $\hat{\beta}_{1i}$ adopted earlier when $T = 1$ is no longer valid as $\text{var}(\hat{\beta}_{1i})$ depends strongly on the number and timing of observed responses which are likely to vary among subjects. More importantly, this ANCOVA approach may be subject to biased inference for γ_1 if the missing data mechanism is other than missing completely at random (Little and Rubin, 1987), e.g. if the length of non-missing period depends on the true but unobserved rate of change, β_{1i} (e.g. Wu and Carroll, 1988).

Table 2. Inference for γ_1 based on multiple post- randomization responses for the schizophrenic trial

Approach	Model 1*	Model 2 *
Full likelihood		
Random intercept	-0.98	-1.13
only	(0.34)	0.33)
Random intercept &	-1.22	-1.34
slope	(0.48)	(0.47)
Conditional likelihood	-1.13	-1.13
	(0.45)	(0.45)
ANCOVA method		
	Without y_{i0}	With y_{i0}
	-1.13	-0.79
	(0.45)	(0.44)

* Model 1: baseline PANSS response is assumed to be different between two groups
Model 2: baseline PANSS response is assumed to be the same between two groups.

For illustration, we return to the schizophrenia example where $T = 5$ with $t_1 = 1, t_2 = 2, t_3 = 4, t_4 = 6$ and $t_5 = 8$ weeks. Figure 3 shows the individual traces of PANSS over the eight week period. It is clear in this display that many persons have missing data. Indeed, only 55 and 44 of 85 subjects from the treated and control group, respectively, were available for the measurement of PANSS in the eighth week. Also note that no subject returned for a visit if he missed the previously scheduled visit, a pattern which we will call "dropping out." Let T_i be the number of the last visit. Table 2 gives estimates of γ_1 and their standard error estimates for three approaches. For the full likelihood approach, results vary according to whether the random slope b_{1i} is allowed and whether z_i is included for baseline PANSS adjustment (Model 1 versus Model 2). With both random

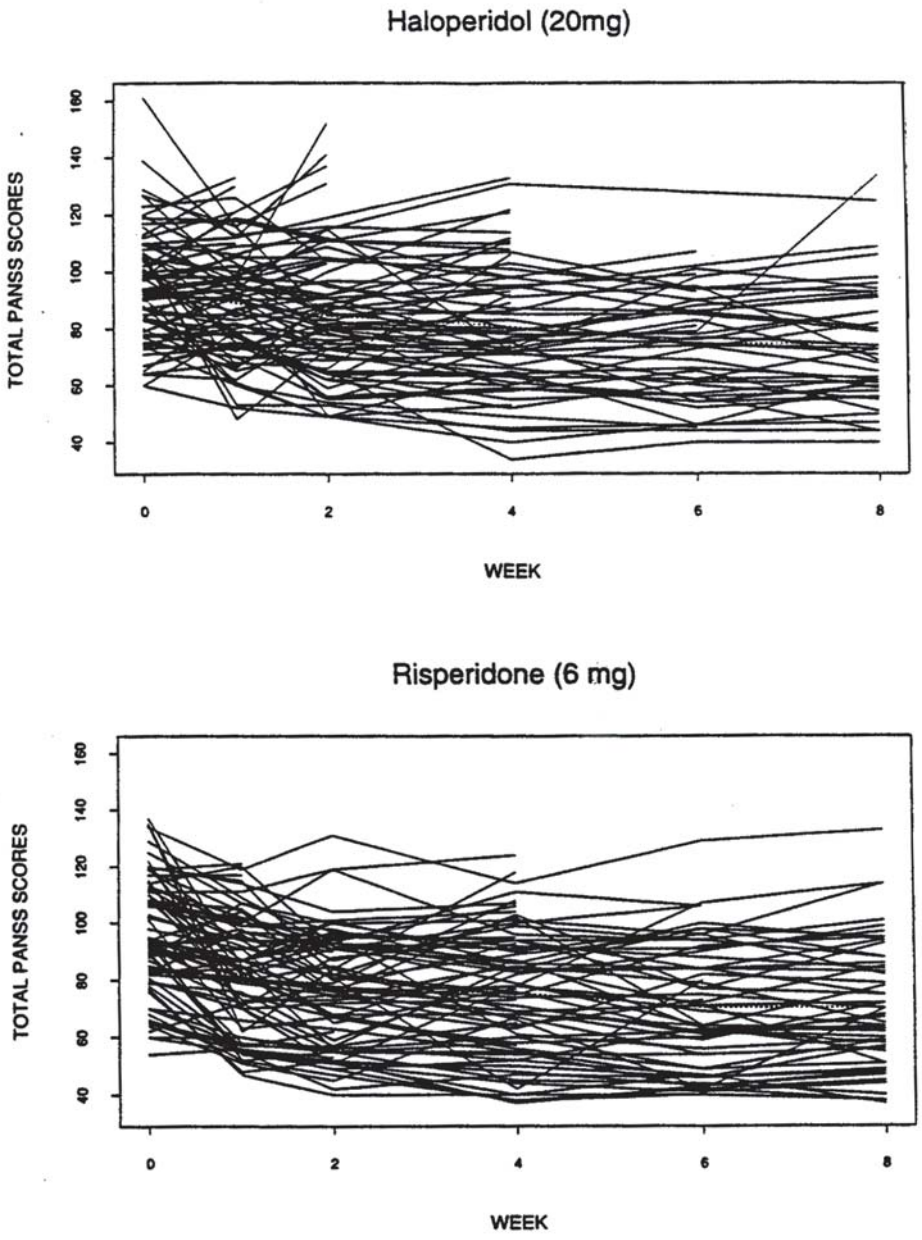

Figure 3: PANSS scores from baseline to eight weeks after randomization; for repeated visits of each patient as connected through line segments

intercepts and slopes to account for natural heterogeneity among subjects, γ_1 is estimated as -1.22 (s.e. $= 0.48$) for Model 1 and as -1.34 (s.e. $= 0.47$) for Model 2. As in §3.1, this example demonstrates that the gain in efficiency for γ_1 estimation is negligible if we assume the baseline PANSS are comparable between the two groups. The conditional likelihood approach based on (10) where T is replaced by T_i to account for dropout gives -1.14 (e.g. $= 0.45$) as the CMLE for γ_1.

As expected, the ANCOVA approach gives identical results to the conditional likelihood approach when z_i is the only covariate. The inclusion of y_{i0} for adjustment purposes does not appear to substantially improve the model fit and the corresponding estimate for $\gamma_1(-0.79 \pm 0.44)$ attenuates. Each of the models leads us to conclude that the new treatment helps on average, to further reduce PANSS at the rate of 1.22 units per week in the eight week period.

4. Statistical Models for Discrete Responses

Primary response variables in health science research are commonly discrete variables. Examples include clinical disease status (normal versus abnormal) or the number of symptoms or clinical events in fixed time intervals. Models for pre-post designs with discrete responses are directly analogous to those for continuous outcomes (1) with the modified assumption that conditional on β_{0i} and β_{1i}, the Y_{ij}'s, $j = 0, 1, \ldots, T$ are independent of each other and follow an exponential family distribution of the form

$$f(y_{ij} \mid \beta_{0i}, \beta_{1i}) = \exp\{\theta_{ij} y_{ij} - a(\theta_{ij}) + c(y_{ij})\} \tag{1*}$$

in which

$$\mathrm{E}(y_{ij} \mid \beta_{0i}, \beta_{1i}) = a'(\theta_{ij}) = g^{-1}(\beta_{0i} + \beta_{1i} t_j),$$

where g is a known "link function." Some common choices of g are the logit link for dichotomous responses and the log link for count responses.

For models involving random intercepts only, i.e. $\mathrm{var}(b_{1i}) = 0, Y_{i+}$ is again the minimal sufficient statistic for b_{0i} when g is the canonical link. Both logit link and the log link are examples of canonical link functions. In this situation, the conditional distribution of Y_i given $Y_{i+} = y_{i+}$ is also from an exponential family with $\Sigma_j y_{ij} t_j$ as the minimal sufficient statistic for θ_2, i.e.

$$f(Y_i \mid y_{i+}; \theta_2) = \exp\{\beta_{1i} \Sigma_j y_{ij} t_j - a^*(\beta_{1i}) + c^*(\Sigma_j y_{ij} t_j, y_{i+})\}, \tag{11}$$

here $\beta_{1i} = \gamma_0 + \gamma_1' z_i$. With dichotomous responses and a logit link, (11) corresponds to conditional logistic regression analysis for highly stratified case-control studies in which n represents the number of strata, y_{1+} and $T - y_{i+}$ the number of "cases" and "controls" in the i^{th}stratum. With count responses and a log link, (11) is proportional to the probability function for a multinomial distribution in which y_{ij} represents the number of trials out of a total y_{i+} that fall into the j^{th}"cell" $j = 0, 1, \ldots, T$. One practical implication of the above observations is that existing

statistical software for analyzing data from stratified case-control studies and for fitting log-linear models to contingency tables is directly applicable. A major concern of this approach is lack of fit for models where $\text{var}(b_{1i})$ has been assumed to be zero. This represents, as in the continuous response case, a major limitation of the conditional likelihood approach since $(Y_{i+}, \Sigma_j t_j Y_{ij})$ are the minimal sufficient statistics for (b_{0i}, b_{1i}) and through conditioning, there is no information left for θ_2 and, in particular, for γ_1.

The conditional distribution in (11) does, however, lend itself to the possibility of extending the ANCOVA approach to discrete responses. Let $\hat{\beta}_{1i}$ be the estimate which maximizes the conditional distribution, $f(Y_i \mid y_{1+}; \beta_{1i})$. The ANCOVA approach suggests that one regress $\hat{\beta}_{1i}$ on z_i with/without y_{i0} adjustment. The problems with this approach is that for non-identity link function, $\hat{\beta}_{1i}$ is no longer an unbiased estimator of β_{1i}. In fact, $\hat{\beta}_{1i}$ may not exist. An example for dichotomous response is when $y_{1+} = T$ or 0, i.e. all the Y_{ij}'s are identical and equal to one (or zero).

This leaves the full likelihood approach as the legitimate alternative for discrete responses allowing both random intercepts and slopes. Under assumption (1*) and (2)-(4) one has

$$L(\theta_1, \theta_2, \theta_3) \propto \prod_{i=1}^{n} \int f(y_i \mid b_{0i}, b_{1i}; \theta_1, \theta_2) f(b_{0i}, b_{1i}; \theta_3) db_{0i} db_{1i}.$$

Note that in maximizing the full likelihood, there is no restriction that the link function be the canonical one. However, the calculation of L involves integrals which must be evaluated numerically even if the bivariate normal distribution is assumed for (b_{0i}, b_{1i}). Several estimating methods have been suggested including Markov Chain Monte Carlo (e.g. Zeger and Karim, 1991) and a penalized likelihood method (Stiratelli, Laird and Ware, 1984; Breslow and Clayton, 1993).

We illustrate the application of conditional and full likelihood approaches using the PANSS example introduced above. A binary response was created from the symptom score as follows: $y_{ij} = 1$ for PANSS ≥ 80; $= 0$ otherwise. Figure 4 displays the prevalence of positive responses against time for the two treatment groups. The first model has a week (γ_0) and week by treatment effect (γ_1) with random intercept only. Table 3 gives the point estimates, standard errors and t-statistics for γ_1 fit by full and conditional likelihood. The second model allows for a random slope as well as intercept and can therefore be fit by full likelihood only. The full likelihood estimates were obtained using an appropriate method due to Stiratelli, Laird and Ware (1984) reviewed by Breslow and Clayton (1993) and implemented in SAS on the GLMMIX macro. Note first that the full likelihood estimate of γ_1 in model 1 (random intercept only) is 25% smaller in absolute value than the conditional likelihood estimate. This ordering is expected because the latter can be approximated by a full-likelihood method with random intercept variance converging to infinity and the estimates of γ_1 tend to increase in absolute size as the random intercept variance increases. Second, note the substantial change in the estimate of treatment effect (γ_1) when both the intercepts and slopes are assumed random while the

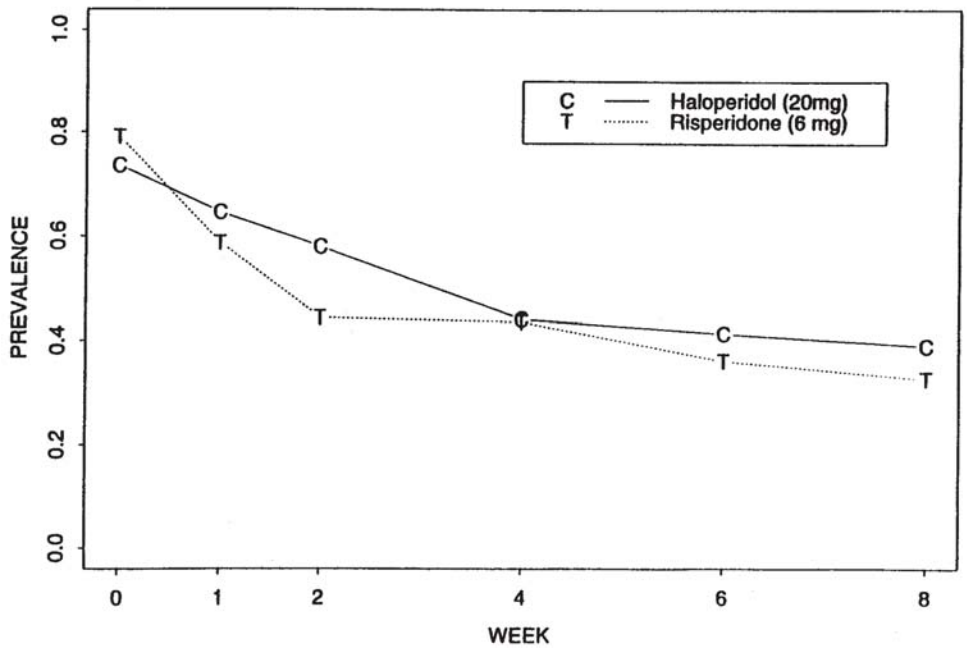

Figure 4: Prevalence of PANSS score ≥ 80 by time and treatment.

t-statistics are largely unchanged. This is also expected and Heagerty (1999) points out that the interpretation and numerical value of coefficients in non-linear random effects models depends critically on the assumed random effects structure. Caution is therefore needed in their application.

Table 3. Inference for the treatment effect (γ_1) in a logistic regression mixed model of PANSS ≥ 80 indicator on time and treatment-by-time

Estimation Approach	$\hat{\gamma}_1$	$\widehat{\text{s.e.}}_{\hat{\gamma}_1}$	t
Full likelihood			
Random intercept only	-.179	.067	-2.69
Random intercept and slope	-.463	.224	-2.06
Conditional likelihood			
Random intercept only	-.237	.100	-2.37

5. Discussion

In this paper, we have reviewed statistical models for the analyses of data collected in clinical trials with the pre-post design. With linear models for continuous responses, we have shown that the traditional ANCOVA estimate obtained by regressing the change (post-pre) on the pre-value and on the treatment group indicator is identical to the estimate obtained from a repeated measures analysis in which both the pre and post-values are considered responses and regressed on a pre-post indicator and pre-post by treatment group interaction. While the ANCOVA formulation is limited to a single post-randomization value and to linear models, the repeated measures, that is longitudinal data analysis (LDA) approach extends naturally to multiple post-values and to discrete data models as well. We have illustrated application of LDA to multiple time points and to binary data using the PANSS schizophrenia measure.

This paper has focused on the LDA approach to data for pre-post designs but has not considered appropriate modelling of missing data. The likelihood-based inferences presented in Tables 1-3 will be valid if the missing data mechanism is ignorable (Little and Rubin,1987). Method for non-ignorable missingness mechanism (Little, 1993; Diggle and Kenward, 1994; Wu and Carroll, 1988) can be directly applied given the LDA approach endorsed here.

The analysis of the binary PANSS response using random intercept and random intercept/slope logistic models shows how sensitive the treatment effect estimate can be to the assumed form of the model, in particular which predictors are assumed to be heterogeneous across subjects. In a linear model, the interpretation of fixed effects is invariant to the specification of the random effects model. In logistic and other non-linear models, it can be highly sensitive.

An alternative approach with the binary response is the "marginal" model (Diggle, Liang and Zeger, 1994) in which logit $\mathrm{E}(y_{ij}|x_{ij})$ and the log odds ratio for (y_{ij}, y_{ik}) are simultaneously modelled. With a marginal model, the interpretation of the mean parameters is invariant with respect to assumptions about the association among repeated observations on a person. Heagerty and Zeger (1998) give a recent review of marginal models for ordinal data.

Acknowledgements. The authors acknowledge that Drs. Alice Hsuan and Peter Quyang from Janssen kindly provided the schizophrenic data set.

References

Allison, P.D. (1990): Change scores as dependent variables in regression analysis. *Sociological Method and Research*, **20**, 93-114.

Breslow, N.E. and Ccayton, D.G. (1993): Approximate inference in generalized linear mixed models. *Jour. Amer. Statist. Assoc.*, **88**, 9-25.

Chouinard, G., Jones, B., Reminton, G.. (1993): A Canadian multicenter placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenic patients. *Jour. Clinical Psychopharmacology*, **13**, 25-40.

DIGGLE, P.J. AND KENWOOD, M.G. (1994): Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49-94.

DIGGLE, P.J., LIANG, K.Y. AND ZEGER, S.L. (1994): *Analysis of Longitudinal Data*, Oxford University Press, New York.

HEAGERTY, P.J. (1999): Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688-698.

HEAGERTY P.J., AND ZEGER, S.L. (1998): LORelogram: a regression approach to exploring dependence in longitudinal categorical responses. *Jour. Amer. Statist. Assoc.*, **93**, 150-162.

LAIRD, N.M. AND WANG, F. (1990): Estimating rates of change in randomized clinical trials. *Controlled Clinical Trials*, **11**, 405-419.

LETTLE, R.J.A. (1993): Pattern-mixture models for multivariate incomplete data. *Jour. Amer. Statist. Assoc.*, **88**, 125-134.

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

MARDER, S.R. AND MEIBACH, R.C. (1994): Risperidone in the treatment of schizophrenia. *American Journal of Psychiatry*, **151**, 825-835.

STANEK, E.J. (1988): Choosing a pre test - post test analysis. *American Statistician*, **42**, 178-183.

STIRATELLI, R., LAIRD, N. AND WARE, J.H. (1984): Random effects models for serial observations with binary responses. *Biometrics*, **40**, 961-971.

WU, M.C. AND CARRPLL, R.J. (1988): Estimation and comparison of changes in the presence of informative light censoring: conditional linear model. *Biometrics*, **44**, 175-188.

ZEGER, S.L. AND KARIM, M.R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Jour. Amer. Statist. Assoc.*, **86**, 79-86.

KUNG-YEE LIANG AND SCOTT L. ZEGER
DEPARTMENT OF BIOSTATISTICS
JOHNS HOPKINS UNIVERSITY
615 N. WOLFE STREET
BALTIMORE MD 21205-2179
U.S.A.
E-mails: kyling@jhsph.edu, szeger@jhsph.edu