Journal of Data Science **9**(2011), 221-241

# Quantifying Treatment Effects When Flexibly Modeling Individual Change in a Nonlinear Mixed Effects Model

Robert J. Gallop[1], Sona Dimidjian[2],
David C. Atkins[3] and Vito Muggeo[4]
[1] *West Chester University,* [2] *University of Colorado at Boulder,*
[3] *University of Washington and* [4] *University of Palermo*

*Abstract*: A core task in analyzing randomized clinical trials based on longitudinal data is to find the best way to describe the change over time for each treatment arm. We review the implementation and estimation of a flexible piecewise Hierarchical Linear Model (HLM) to model change over time. The flexible piecewise HLM consists of two phases with differing rates of change. The breakpoints between these two phases, as well as the rates of change per phase are allowed to vary between treatment groups as well as individuals. While this approach may provide better model fit, how to quantify treatment differences over the longitudinal period is not clear. In this paper, we develop a procedure for summarizing the longitudinal data for the flexible piecewise HLM on the lines of Cook *et al.* (2004). We focus on quantifying the overall treatment efficacy using the area under the curve (AUC) of the individual flexible piecewise HLM models. Methods are illustrated through data from a placebo-controlled trial in the treatment of depression comparing psychotherapy and pharmacotherapy.

*Key words*: Area under the curve (AUC), breakpoint, Hierarchical Linear model, random effects, subject-specific estimates.

## 1. Introduction

Randomized controlled trials (RCTs) have long focused on treatment effects over time, that is, longitudinal data. Hierarchical linear models (HLM) provided an important statistical advance in clinical trial methodology; however, the typical use of HLM assumes a linear trend across time for each person, which may not be valid for modeling clinical change trajectories. Many extensions of the HLM model to accommodate non-linear changes of time have been implemented in recent years such as log-linear change (Gibbons *et al.*, 1993), piecewise linear models (Verbeke and Molenberghs, 2000), and quadratic change (Cook *et al.*,

2004). Cudeck and Klebe (2002) and Cudeck and Harring (2007) presented a multiphase mixed-effects model, which we refer to as a flexible HLM model because it consists of two phases with differing rates of change for each treatment group. The breakpoints between these two phases, as well as the rates of change per phase are allowed to vary between treatment groups as well as individuals. Despite the complexity in modeling change over time in a multiple treatment arms RCT, an aim of the RCT is to quantify how the treatment arms differ over the longitudinal period. Contrasting the individual components of the flexible HLM model will not answer this aim. The main objective of this paper is to summarize treatment arm differences over the entire longitudinal data. The present article extends the Area under the Curve (AUC) calculations of Cook *et al.* (2004) to serve as an index to summarize change over the longitudinal period.

In the following sections, we present a detailed discussion of how to quantify treatment effects with flexible piecewise models illustrating the implementation using longitudinal data from a recent placebo controlled RCT comparing psychotherapy (PSY) and pharmacotherapy (ADM) in the treatment of adults with major depression (Dimidjian *et al.*, 2006). We compare two active treatment arms, psychotherapy (PSY) and pharmacotherapy (ADM), with pill-placebo controls (PBO). Each patient completed multiple self report measures of depressive severity (Beck Depression Index; BDI; Beck, Steer, and Brown, 1989). Repeated measures consisted of two primary assessment points (pre-treatment, endpoint at Week 8) and at each treatment session.

In Section 2, we discuss the advances offered by HLM and the limitations for modeling longitudinal data in many clinical trials. We then discuss the implementation and estimation of a flexible piecewise HLM, consisting of two treatment phases with differing rates of change for each treatment group: an early rapid phase of change followed by a phase of reduced change, for which the breakpoint between these two phases is allowed to vary across both treatment groups and individuals (Section 3). In Section 4, we illustrate the application of the flexible piecewise model to this recent RCT. In Section 5, we discuss strategies for assessing treatment efficacy starting with a generic example followed by a specific example of application to this recent RCT. Sensitivity analyses and assessment of fit are in Section 6. Some concluding remarks, limitations, and discussion of potential areas of future research are made in Section 7.

## 2. Standard HLM for Longitudinal Data Analysis

Longitudinal data are one type of correlated data, though data that have a nested or hierarchical structure are common in a wide variety of disciplines (e.g., students nested within classrooms, family members nested within families). Similar models for the analysis of longitudinal data have been developed across

disparate disciplines, under a variety of names: random coefficient models, hierarchical linear models, and multilevel linear models (Goldstein, 1987; Bryk and Raudenbush, 1996; Raudenbush and Bryk, 2002).

Within statistics, Laird and Ware (1982) and Sternio, Weisberg and Bryk (1983), working independently, proposed essentially identical approaches to the analysis of repeated measures data, both using the EM algorithm. In this approach, each person's growth curve is characterized by a set of person-specific parameters, i.e., random effects. For example, in a linear growth model, the parameters are a mean and rate of growth. However, these parameters are themselves viewed as randomly sampled from a population of individuals. Hence, the model may be viewed as having two levels: a within-person level, which is referred to as the level 1 portion of the model, and a between-person level, which is referred to as the level 2 portion of the model. In the present paper, we refer to these models as hierarchical linear models (HLMs; Raudenbush and Bryk, 2002).

One advantage of HLM is that it can incorporate imbalanced longitudinal data and does not require assessments of outcome at common discrete points of time. Even the most extreme case, in which all patients had unique assessment times, can be accommodated with this framework. A critical issue in fitting HLM to longitudinal data is to accurately quantify change as a function of time. The starting point to quantify change usually begins by specifying the level 1 equation that characterizes change over time:

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \tag{1}$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$. In this equation $y_{ij}$ represents the dependent variable measured for individual $i$ at the $j^{th}$ assessment point, $\beta_{0i}$ represents the intercept or the dependent variable for individual $i$ at time 0 (usually the pre-treatment assessment), $\beta_{1i}$ represents the linear rate of growth for individual $i$ across each time point, $t_{ij}$ represents the exact time for the $j^{th}$ assessment point of individual $i$, and $e_{ij}$ is the residual or error term indicating the deviation of each individual's score from their own modeled line. The level 1 parameters in HLM become dependent variables in level 2 of the model. In this way the parameter estimates related to "time" at level 1 are nested within the person at level 2. With RCT data the typical focus is on the use of HLM in determining effects between treatment conditions in rate of change over time. In our example, treatment condition would refer to the $i^{th}$ patient's randomization assignment to PSY, ADM, or PBO, which we represent as $TX_i$. The level 2 equations as a function of treatment condition, represented as $TX$, are as follows:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(TX_i) + u_{0i} \tag{2}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(TX_i) + u_{1i} \tag{3}$$

where

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}\right).$$

If one does not expect the outcome to change across time in a linear fashion, or the data does not support such change, then a refinement of the model is required. Linear time assumes the rate of change is constant across the entire longitudinal period under investigation; however, much longitudinal change is not constant over time. Our goal should be to model the data and not place the data into standard models when the standard model does not apply. When change is not linear, we must accommodate the data accordingly, which starts with flexibly modeling change. At the same time, we must balance the goals of the analysis, which in our case is contrasting the treatment effects, with the complexity of the model. We must still be able to describe and interpret the results with respect to our goals of the analysis, regardless of the model complexity.

## 3. Flexible Piecewise Model

It is quite common in an RCT to see an initial phase of substantial change followed by a second phase with reduced change. Keller *et al.* (2000) in a 12 week study found two phases of change over time with the distinct phases corresponding to change from baseline to week 4 and change from week 4 to week 12. This type of change over time structure is modeled as a piece-wise linear model, where rates of change are allowed to differ between the two phases. In essence, two rates of change connected at a point of change referred to as a breakpoint, are estimated for each subject. The location of this breakpoint is at times determined by study design features such as location of mid-point of the study, change in frequency of assessment, change in medication regimen, or progressing from active treatment phase to a follow-up phase. When we considered individual trajectories based on the data from the Dimidjian *et al.* (2006) study, profiles showing this piecewise structure become more apparent, as seen in Figure 1.

The four patients' trajectories, with individual depression scores measured by the Beck Depression Index; BDI (Beck, Steer, and Brown, 1989) on the y-axis and week since randomization on the $x$-axis, illustrate the following: (a) piecewise profile, (b) the early rate of change between the four patients is different, (c) the later rate of change between the four patients is different, and (d) the location of the break between early and late change is different for each in contrast to typical piecewise linear models, which consider a common breakpoint for all treatments. Outside of RCTs this dual-phase pattern of change appears common across many areas of psychological research (Cudeck and Klebe, 2002; Cudeck and Harring,

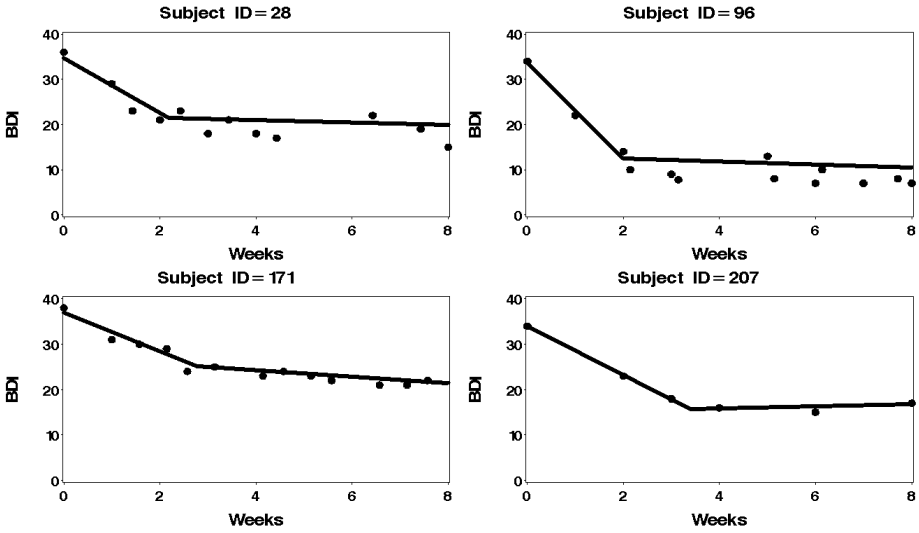


Figure 1: Individual piecewise linear change for four patients

2007). From a clinical perspective, this differential breakpoint implies that the significant impact of a given treatment may be more effective in a smaller time window compared to the other treatments.

Similar to standard HLM models where subject-specific effects are introduced to allow subject heterogeneity, flexible HLM introduces a random effect due to the location of the breakpoint to allow subjects' change of phase to deviate from the overall average location for their treatment group. Following Cudeck and Klebe (2002) the level 1 equation is:

$$y_{ij} = \beta_{0i} + \beta_{1i} \min(t_{ij}, \tau_i) + \beta_{2i} \max(0, t_{ij} - \tau_i) + e_{ij}, \tag{4}$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$. In this equation $y_{ij}$ represents the dependent variable measured for individual $i$ at $j^{th}$ assessment acquired at time $t_{ij}$. Unknown parameters to be estimated are: $\beta_{0i}$ which represents the intercept or the dependent variable for individual $i$ at time 0; $\beta_{1i}$ which represents the linear rate of growth for individual $i$ across each time point of the early phase of change; $\beta_{2i}$ which represents the additive component for the linear rate of growth for individual $i$ across each time point of the later phase of change; and $\tau_i$ which represents the subject-specific breakpoint between early and late phase of change. Equation (4) is referred to as a connected piecewise linear model, where the last assessment during the early phase of change serves as the starting point for the later phase of change, with $e_{ij}$ as the residual or error term indicating the deviation of each individual's score from their individual connected piecewise linear model. With $\tau_i$ unknown, this equation is nonlinear due to the random breakpoint being included

as a multiplicative term to the legs of the piecewise model. As in the standard HLM, the level 1 parameters become dependent variables in level 2 of the model. The level 2 equations as a function of treatment condition are as follows:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(TX)_i + u_{0i} \tag{5}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}(TX)_i + u_{1i} \tag{6}$$
$$\beta_{2i} = \gamma_{20} + \gamma_{21}(TX)_i + u_{2i} \tag{7}$$
$$\tau_i = \gamma_{30} + \gamma_{31}(TX)_i + u_{3i} \tag{8}$$

where

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \\ u_{3i} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} & \tau_{03} \\ \tau_{01} & \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{02} & \tau_{12} & \tau_{22} & \tau_{23} \\ \tau_{03} & \tau_{13} & \tau_{23} & \tau_{33} \end{bmatrix} \right).$$

With regards to estimation of the parameter estimates, standard likelihood-based inferences is complicated by the fact the breakpoint is random with the log-likelihood being just piecewise differentiable at this breakpoint, resulting in a situation where classical regularity conditions are not met (Feder, 1975; Kuchenhoff and Ulm, 1997; and Muggeo, 2003). Several methods have been proposed to approximate parameter estimation: grid search, smoothing, a mixture approach, or Bayesian methods.

Grid search estimation, as its name indicates, is performed by a grid search over the parameter space, (Hawkins, 1976; Ertel and Foulkes, 1976; Ulm, 1991; Rigby, 1992; Stastinopoulus and Rigby, 1992). In smoothing (Bacon and Watts, 1971; Seber and Wild, 1989; Morrell *et al.*, 1995), the point of non-differentiability at the breakpoint is smoothed through an arithmetic transformation in the vicinity of the breakpoint. Common applied transformations are the hyperbolic tangent function or a rational function smoothers as discussed by Dietz and Hadeler (1988) for the minimum and maximum functions. EM algorithmic approach and iterative procedures have been suggested by Scott *et al.* (2004) and Muggeo (2003), respectively, which have shown to produce convergence with robust estimates. Dominicus *et al.* (2006, 2008) used Gibbs sampling under a Bayesian perspective for parameter estimation.

We will proceed with the analytic approximation of the likelihood function. The rational function smoother of Dietz and Hadeler (1988) was implemented to correct the piecewise differentiability at the breakpoints in equation (4), hence correcting the classical regularity conditions. Initial parameter estimation was made using the first-order linear (FIRO) approximation method of Beal and Sheiner (1982). Dominicus *et al.* (2006) showed slight bias with the FIRO

method when compared to Bayesian estimation. With the random effects following a multivariate normal distribution, a more accurate numerical integration of the likelihood function is achieved with Gauss-Hermite quadrature (Davidian and Giltinan, 2003; Wang, 2007). Accuracy of the approximation was based on increasing the number of quadrature points from 1 to 30 points.

All modeling operates under the philosophy of parsimony. While we are proposing a flexible piecewise model, we must inspect whether there is truly two different phases of change. Without two phases of change, the level 1 model will be simplified to equation (1), by removing both the second leg of change and the random breakpoint. The need for the two phases of change will be based on the approach described by Piepho and Ogutu (2003), which assesses the need of the breakpoint through the null hypothesis that the slope prior and post the breakpoint are equal. This flexible piecewise model and the model diagnostics can be fit through many software programs for non-linear mixed effects models.

## 4. Application to Treatment of Depression Study

To illustrate the flexible HLM methods proposed, we utilize data from a recent placebo controlled RCT comparing psychotherapy (PSY), pharmacotherapy (ADM), with pill-placebo controls (PBO) described in section 1. The analyses consist of 124 patients who met criteria for moderate to high depression severity at the start of the study. Of these patients, 48 were randomly assigned to PSY, 49 to ADM, and 27 to PBO (Dimidjian *et al.*, 2006). The maximum number of sessions allowed per protocol varied by treatment condition; thus, the potential number of outcomes and time between repeated assessments varies by treatment. Specifically, ADM patients were permitted up to 6 sessions; PBO patients were permitted up to 6 sessions; and PSY patients were permitted up to 16 sessions. Further details of the study can be found in Dimidjian *et al.* (2006).

We fit the flexible piecewise model per equations (4)-(8). Covariance structure was based on comparison of the -2Log-Likelihood, Akaike information criterion (AIC), and the Akaike information corrected criterion (AICC) of nested models indicating the need for variance components for all subject-specific parameters, covariance between two slope components, and covariance between the late slope and the breakpoint. The need for breakpoints per group was verified through the approach of Piepho and Ogutu (2003) and Kim *et al.* (2004). Statistical tests were based on a t-distribution with degrees of freedom based on the sample under analysis, which is 124 subjects, with a loss of 4 degrees of freedom based on the subject-specific parameter per equation (4). Results indicated the need for the breakpoint for each treatment: PSY ($t(120) = 2.31$, $p = 0.02$); ADM ($t(120) = 4.33$, $p < 0.0001$; and PBO ($t(120) = 4.64$, $p < 0.01$). Parameter

estimates are illustrated in Table 1.

Table 1: Parameter estimates for piecwise Model

| | **Fixed Effects** | | |
|---|---|---|---|
| | **PSY** | **ADM** | **PBO** |
| **Intercept** | 35.08 (0.99) | 35.82(0.98) | 34.19 (1.33) |
| **Slope during leg 1** | -2.62 (0.36) | -3.61 (0.30) | -3.47 (0.48) |
| **Slope during leg 2** | -1.48 (0.21) | -1.44 (0.30) | -0.12 (0.37) |
| **Breakpoint** | 2.83 (0.26) | 3.62 (0.41) | 2.91 (0.40) |
| | **Random Effects** | | |
| $\tau_{00}$ | 42.12 (5.87) | | |
| $\tau_{11}$ | 11.40 (1.87) | | |
| $\tau_{22}$ | 2.90 (0.53) | | |
| $\tau_{33}$ | 0.75 (0.24) | | |
| $\tau_{12}$ | -3.92 (0.85) | | |
| $\tau_{23}$ | 0.63 (0.27) | | |
| $\sigma^2$ | 17.93 (0.89) | | |

Note: Standard error estimates are in parentheses.

As seen in table 1, ADM and PBO have a more rapid reduction in depressive symptoms during the early phase of therapy compared to PSY. Although, performing statistical contrasts indicates a statistically significant difference in the rate of change during the first leg for ADM compared to PSY only: ADM compared to PSY ($t(120) = 2.49$, $p = 0.014$); ADM compared to PBO ($t(120) = 0.26$, $p = 0.80$) and PSY compared to PBO ($t(120) = 1.57$, $p = 0.12$). During the late phase of treatment, PBO has the slowest rate of reduction in symptoms. Statistical contrasts indicates the rate of change during the second leg is significantly different for PSY compared to PBO ($t(120) = 3.17$, $p = 0.002$) and ADM compared to PBO ($t(120) = 2.85$, $p = 0.005$). Focusing on the breakpoints, we see ADM group on-average stays on their "early phase of change" for half a week to a week longer compared to both PBO and PSY. Statistical contrasts indicate no statistically significant difference in breakpoints between the three groups. The key question is how does each portion of the flexible HLM model contribute to the overall treatment contrast during the longitudinal period?

## 5. Quantifying Treatment Efficacy

Based on the parameter estimates of Table 1, the on-average trajectories for average depression score, the y-axis, as measured by the BDI, versus weeks since randomization, the x-axis, are illustrated in Figure 2.

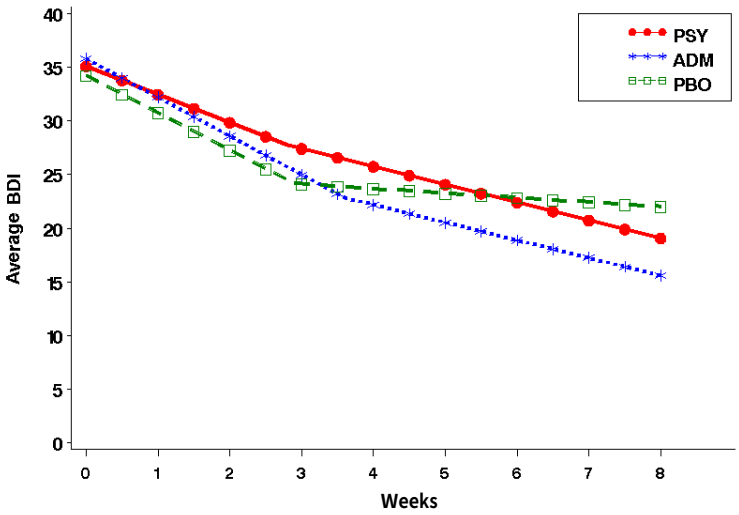


Figure 2: On-average piecewise profiles for the Treatment of Depression Study

How patients change over time and the on-average behavior between groups is different. With a linear model for time, the interaction between treatment and time provides a single, accurate assessment of how the treatments differ. However, with the flexible piecewise model just presented, there are a variety of treatment effects possible. To compare treatment groups, contrasts between the early slopes only indicate how the treatments differ during their early phase of treatment. As we see in Figure 2, all treatments experience a rapid phase of reduction followed by a period of reduced reduction. The amount of time on the early phase of change varies between the treatments, and the magnitude of change per phase varies between the treatments. It is unfair to contrast the treatments based on slope for either leg of the piecewise model, since the amount of time on either leg, directly linked to the breakpoint, is different between treatments. We need an approach that simultaneously considers the entire trajectory, the starting point, and the breakpoint, to derive an overall approach to contrast the groups, while not focusing on specific portions of the flexible piecewise model.

Response feature analyses reduce complex longitudinal data to simple summary indices to allow for hypothesis testing and statistical contrasts (Everitt, 1995). To answer our hypotheses of differential treatment effect on depression across the study, we compute the area the under the curve (AUC) per person (Matthews *et al.*, 1990). The AUC is a familiar index seen in logistic regression model, which summarize the discrimination of a predictive model in the analysis of a binary outcome. However, the present derivation of AUC is different than this common application in logistic regression. We use the AUC to summarize an individual's trajectory, consisting of repeated measures on a continuous outcome.

This derivation of AUC is seen quite often in the analysis of pharmacokinetic data, which reflects the total exposure of an individual subject to the administered medication during a set dosing period (Nguyen and Amaratunga, 2001). In our setting, the AUC will represent the total depression severity over the active treatment period. This AUC offers numerous advantages, including that it is easily derived, does not require balanced data, and allows us to compare the overall difference between groups. The disadvantage of the AUC is we lose information about the time process. With this summary index, which reduces the longitudinal data to one summary measure, we can perform basic cross-sectional analyses such as analysis of variance (ANOVA) or analysis of covariance (ANCOVA) to contrast the treatment groups on their performance over the longitudinal period. Thus, the combination of the flexible piecewise model coupled with the AUC contrast will provide a clear picture of how treatment groups change on-average over time, hence no loss of information about the time process, while also providing a means for assessing the treatment differences over the longitudinal period.

Focusing on equation (4), we derive the individual AUC to characterize the total depression score through the 8 weeks of active treatment. This is defined as the integral of the predicted depression curve over the 8 weeks of active treatment. For the patient, the individual AUC is given by:

$$AUC_i = \int_0^8 \left[\beta_{0i} + \beta_{1i} \min(t, \tau_i) + \beta_{2i} \max(t - \tau_i, 0)\right] dt, \tag{9}$$

where $t$ is representing time in weeks. A similar approach was taken by Cook *et al.* (2004) in their inspection of children's blood pressure from ages 5 to 14, which was described through a quadratic random-effects model over unequally spaced repeated measures. The solution of the integral in equation (9) is:

$$AUC_i = 8\beta_{0i} + 8\beta_{1i}\tau_i + \frac{(8 - \tau_i)^2}{2}\beta_{2i} - \beta_{1i}\frac{\tau_i^2}{2}. \tag{10}$$

With the summary individual AUC measure, which is derived based on the subject specific intercept, slope on the early phase of change, slope on the late phase of change, the individual breakpoint, and the elapsed time of the treatment (8 weeks), we can contrast the groups. These estimates indicate deviations for each patient from their respective treatment's average effects. These estimates are termed Best Linear Unbiased Prediction (BLUP) estimates. BLUP estimates are: linear in the sense that they are linear functions of the data, unbiased in the sense that the average value of the estimates is equal to the average value of the quantity being estimated, best in the sense that they have the minimum mean squared error within the class of linear unbiased estimators, and prediction estimates to distinguish them from estimation of the random effects (Robinson,

1991). Thus rather than consider these specific effects as nuisance parameters, with a central focus on contrasts of treatments, these subject specific estimates per equation (4) are used as the "individual expected AUC" for a given patient. We will test for differences by treatment using this summary measure as the primary outcome in a response feature analysis. Application of this approach on our data is discussed in the next section.

## 6. Application of the AUC to the Treatment of Depression Study

From Figure 2, we see comparable rates of change for ADM and PBO during the early phase and comparable rates of change for PSY and ADM during the late phase, but the key question we wish to answer is how do the treatment specific piecewise trajectories contribute to the overall treatment contrast during the entire longitudinal period, where we're most interested in seeing superiority to the control group, PBO. To answer this we proceed with the individual AUC derivation per equation (10).

We have AUC estimates of 204.58 for PSY, 192.95 for ADM, and 222.20 for PBO. So over the 8 week placebo control period, on-average we have lower symptoms for ADM and PSY relative to PBO, which were statistically contrasted using ANCOVA. The ANCOVA model adjusted for baseline severity, for, as seen in table 1, the intercepts vary by group. While the difference at baseline is not statistically significant, which is attributable to the integrity of the treatment randomization, we still need to adjust the individual AUC analysis for the observed pre-treatment difference. In addition, we included gender as a covariate. As discussed in Dimidjian *et al.* (2006), the psychotherapy group had fewer women. The overall treatment contrast of group AUCs resulted in a marginally significant treatment effect ($F(2, 119) = 2.94$, $p = 0.057$). The specific treatment contrast with the placebo control group are $t(119) = 1.39$, $p = 0.17$ for PSY compared to PBO and $t(119) = 2.41$, $p = 0.02$ for ADM compared to PBO. We see a significant effect for ADM compared to PBO; therefore, during the 8-week placebo-control period, ADM results in patients with less severity on-average compared to PBO, whereas, based on the available data, we do not have sufficient evidence indicating on-average superiority of PSY compared to PBO.

## 7. Sensitivity Analyses

We see from Table 1 that each group goes from a period of rapid reduction to a period of less reduction, one concern is whether the flexible piecewise model truly fits the data best. Focusing on the treatment average breakpoint per Table 1, we see breakpoint ranges within one week of each other. To examine the overall fit of the flexible piecewise HLM to the data, it was contrasted with five alternative

models (level-1 equations):

$$\begin{aligned} y_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + e_{ij} \\ y_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} t_{ij}^2 + e_{ij} \\ y_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} t_{ij}^2 + \beta_{3i} t_{ij}^3 + e_{ij} \\ y_{ij} &= \beta_{0i} + \beta_{1i} \min(t_{ij}, 4) + \beta_{2i} \max(t_{ij} - 4, 0) + e_{ij} \\ y_{ij} &= \beta_{0i} + \beta_{1i} \log(t_{ij} + 1) + e_{ij}. \end{aligned} \tag{11}$$

In equation (11), we consider a series of different functions of time in our level 1 model: linear, quadratic, cubic, piecewise linear model with a common breakpoint at the midpoint (i.e., week 4, paralleling the approach of Keller *et al.*, 2000), and shifted logarithmic model paralleling Gibbons *et al.* (1993). Fit of the data for each model, as well as our proposed flexible piecewise model per equation (4), is summarized in Table 2 based on -2 Log-likelihood functions, the Akaike information criterion (AIC), and the Akaike corrected information criterion (AICC).

Table 2: Comparison to alternative models

| | **-2 Log-Likelihood** | **AIC** | **AICC** |
|---|---|---|---|
| **Linear** | 7777.0 | 7785.0 | 7785.1 |
| **Piecewise Linear common breakpoint** | 7602.4 | 7630.4 | 7630.8 |
| **Quadratic** | 7625.3 | 7639.3 | 7639.4 |
| **Cubic** | 7634.9 | 7648.9 | 7649.0 |
| **Shifted Logarithmic of time** | 7734.0 | 7756.0 | 7787.0 |
| **Flexible model per equation (4)** | 7583.0 | 7617.0 | 7617.6 |

As we see in Table 2, the flexible model proposed in equation (4) yields the best fit indicated by the lower values of each index. Even compared to the piecewise model with a common breakpoint, the addition of the random break is warranted. This is based on the difference of the -2 Log-Likelihood is 19.4, which is significantly different based on a chi-square distribution with 4 degrees of freedom, corresponding to an estimates for each treatment and even a chi-square distribution with 8 degrees of freedom, corresponding to the estimate for each treatment, variance component for the random breakpoint, and covariance estimates for each of the other three components: random intercept, random first slope, and random second slope.

To take the model diagnostic to the patient observed data, we compared the predicted estimates of the flexible piecewise model per equation (4), the predicted estimates of the shifted logarithmic model per equation (11) and the observed BDI scores. The reason we focus on the shifted logarithmic model per equation (11) is because this approach has received applications in many psychological setting, with the most recognized example discussed in Gibbons

*et al.* (1993). Within each subject, we derived the correlation of the observed data versus the subject-specific estimates in the two models. Therefore, for each subject we have a measure quantifying how well the observed data associated with the predicted estimates relate. Across all 124 patients in our sample, for the flexible piecewise model we had a median correlation of 0.861 with an interquartile range of 0.683-0.936. Therefore, we see for 50% of the sample or more we have a strong correlation of the predicted estimate with the observed data per subject with correlations of 0.683 or larger. For the shifted logarithmic model, we have a median correlation of 0.782, with an interquartile range of 0.553 - 0.912. For the shifted logarithmic model, the correlation is not as strong, with correlations of 0.553 or larger for 50% of the sample. Hence, the stronger association of the observed data compared to the estimated data per the flexible model in equation (4) provides evidence of the better fit of the model to the data.

We next extended the goodness of fit approach, to our index of quantifying the treatment effect, the AUC. To do this we derive an observed AUC score per person. Refer to Figure 3.

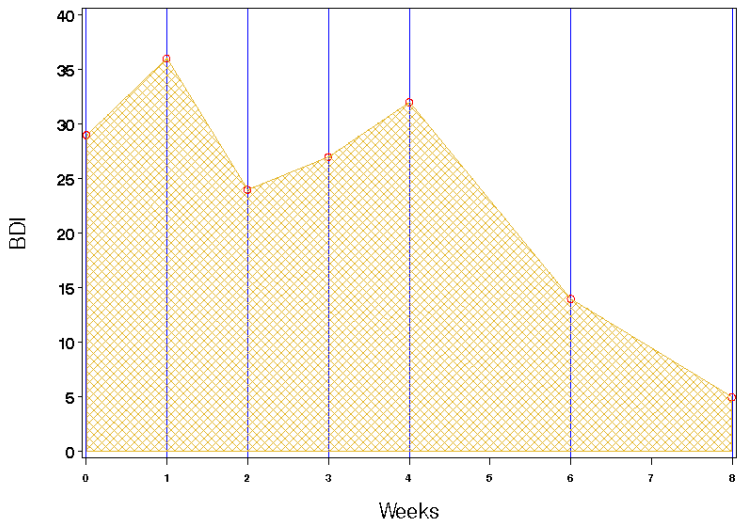


Figure 3: Observed individual AUC

To derive this observed AUC score, we recognize the shape of the profile between sequential points, per Figure 3, is a trapezoid, with the week axis serving as the height, and the previous and current assessment serving as the two bases. The area for each trapezoid is the product of the average base times the height. Derivation of the observed AUC per person is as follows:

$$ObsAUC_i = \sum_{k=1}^{n_i} \left( \frac{y_{i(k-1)} + y_{ik}}{2} \right) \times \Delta_k \tag{12}$$

where for the $i^{th}$ person, there are $n_i$ data points, with $y_{ik}$ as the outcome at the $k^{th}$ assessment, and $\Delta_k$ is the elapsed time between the $k^{th}$ assessment and the previous assessment. The value $y_{i0}$ is the outcome at time 0, the baseline value. While the elapsed duration of treatment was scheduled for 8 weeks, completion deviates from the targeted date at week 8. Therefore, AUC scores per person are pro-rated for 8 weeks. Observed AUC scores per treatment are derived and compared to the individual AUC scores as illustrated in Figure 4.

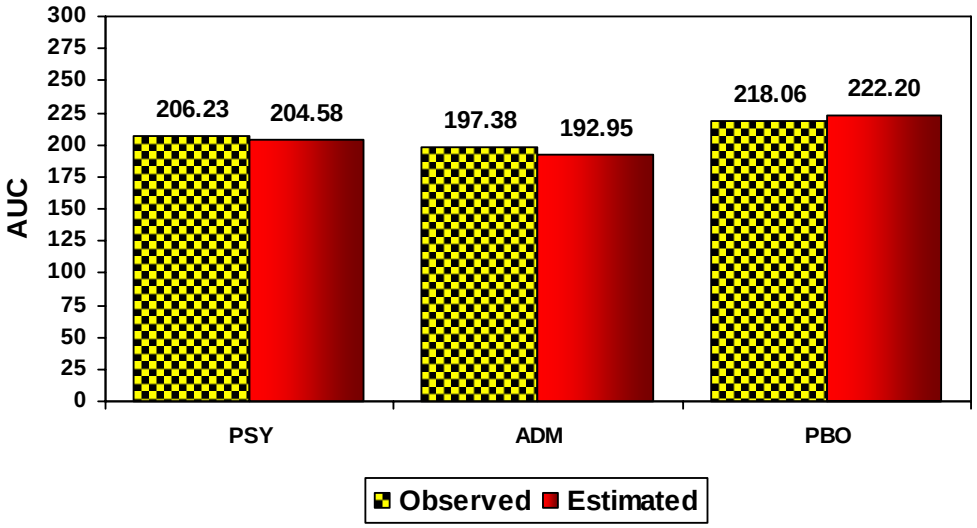


Figure 4: Observed and individual on-average AUC estimates

As illustrated in Figure 4, the model based estimates are all close to the observed estimate.

Thus, these diagnostic approaches all point to the goodness of fit of the model described in equation (4). We therefore have an adequately fitting model to describe the change over time, plus a summary measure, the AUC, to allow us to contrast the overall treatment effect.

## 8. Conclusion

Software algorithms have provided us much flexibility in deriving sophisticated models to mirror the observed data; however, the difficulty in interpreting the overall treatment effect becomes a challenge as the complexity of the model increases. The reported analysis from this depression study, based on the endpoint assessment point, which examined for treatment differences based on an Analysis of Covariance (ANCOVA) covarying baseline found a marginally significant effect for treatment over the 8 weeks ($F(2, 119) = 3.00$, $p = 0.054$) (Dimidjian *et al.*,

2006)., where it is seen that the ADM patients on-average had marginally more rapid reduction compared to PBO patients ($t(119) = -1.84$, $p = 0.07$). The additional data used in the flexible HLM analysis, provides a more descriptive change profile over the 8 week placebo-control period, while the AUC index analysis produces results in similar directions, yet a more powerful result compared to the cross-sectional ANCOVA analysis. Besides being able to summarize the total depression severity over the 8 weeks, we see similar to Cook *et al.* (2004) application, the AUC used all available data per individual and can accommodate unbalanced repeated measurements, while providing a single summary measure per person to perform a response feature analysis to examine the treatment effect over the longitudinal period.

One limitation of the analytical approach is to the sensitivity of HLM structures to the assumption of random effects (Heagerty and Zeger, 2000). With the derivation of the AUC based on the subject-specific estimates from the random terms (random intercept, random first slope, random second slope, and random breakpoint), the assumptions of normality of the random effects must be met to adhere to the warning of Heagerty and Zeger (2000). Based on our data, we saw no deviations of normality for any of the random effects.

Lee and Thompson (2008) discuss relaxing the normality of random effect distributions in the HLM structures, citing that normality assumption can be extremely restrictive. Their work incorporates random effects with distributions from a skew extension of the normal and t-distributions, which may provide better model fit. With the random breakpoint, in this setting, representing time, consideration of alternative distributions for the random term may provide a worthwhile area of future research. We know time, in our application, is non-negative. Therefore consideration of distributions which have non-negative support sets may be appropriate for the random breakpoint. Muggeo *et al.* (under review) has taken this further and has considered no distribution on the random breakpoint, treating it non-parametrically.

For repeated assessments, missing data are inevitable, especially with attrition, but the key issue is whether the inferences are impacted by the presence/absence of the data. As discussed one of the advantages of the AUC is the flexibility in handling unbalanced data. From a statistical inference perspective, the HLM models, which assume that data are missing at random (MAR), are especially robust compared to other longitudinal data analysis procedures such as Generalized Estimating Equations (GEE), which makes a more restrictive assumption, i.e., data missing completely at random (MCAR; Diggle, Liang, and Zeger, 1994; Little, 1994; Hedeker and Gibbons, 1997). MAR means the missing data process is independent of the value of the outcome variable (e.g., depression scores) but can depend on some other observed variable in the study

(e.g., race, gender, and treatment condition). Spritzler *et al.* (2008) discuss impact of missing data to AUC estimation. As far as approaches to accommodate when the data is not MAR is a current limitation of this analytical approach and a possible additional area of future research. Two additional limitations of this application focus on limitations of the BLUP estimates. Distribution theory associated with BLUP estimates is not nearly as well-understood as it is with conventional estimable functions (Littell *et al.*, 1995). The variance of the BLUP estimates may experience shrinkage, since the observed data are shrunk towards the overall average since the prior means of the random effects is zero (Verbeke and Molenberghs, 2000).

As with most algorithms the initial estimates are important in achieving maximization of the likelihood (Powell, 1970). Despite computer speeds, numerical quadrature is computationally burdensome. As illustrated Gauss-Hermite quadrature method coupled with using initial parameter estimates based on the first-order linearization of Beal and Sheiner (1982), we are able to derive parameter estimates. The advantage is that this "approximate likelihood" is available in a closed form and standard errors are obtained from the information matrix (Davidian and Giltinan, 2003). Consistency of parameter estimates was based on increasing the number of quadrature points from 1 to 30, where 1 quadrature point corresponds to a Laplacian method. Laplacian methods are considered a highly accurate method for approximating parameter estimates for nonlinear mixed effects model (Wang, 2007). The accuracy of the approximation increases as the number of quadrature points increases (Davidian and Giltinan, 2007). Our model diagnostics provide evidence of goodness-of-fit. Although in our application we did not fit an unstructured covariance matrix for the random effect. Our examination of nested models with more complex covariance structures compared to more simplistic structures, provided evidence that some covariance terms were going to 0. Therefore, there was no appreciable loss of model fit with the more simplistic covariance structure.

We have presented an application of flexible mixed effects model for estimation of depression severity measured over 8 weeks of active placebo-control treatment period. The attractive feature of the flexible mixed effects model is it describes the different facets of treatment within each group. While the flexible mixed effects model summarized the individual change over time, we derived individual AUC estimates based of the flexible mixed effects model. We saw similar values for the model-based individual AUC estimates and the observed AUC estimates. The AUC index served as the outcome measure in our response feature analysis for contrasting the treatment groups.

**Acknowledgements**

The preparation of this manuscript was funded in part by grant MH55502 (N.S. Jacobson/ D. Dunner). We appreciate the discussion with Jeffrey Harring concerning parameter estimation methods.

## References

Bacon, D.W. and Watts, D.G. (1971). Estimating the transition between two intersecting straight lines, *Biometrika* **58**, 525-534.

Beal, S.L., and Sheiner, L.B. (1982). Estimating Population Pharmacokinetics, *Critical Reviews in Biomedical Engineering* **8**, 195-222.

Beck, A.T., Steer, R.A., and Brown, G.K. (1996). *Manual for the BDI-II.* The Psychological Corporation, San Antonio.

Byrk A., and Raudenbush S. (1996). *Hierarchical Linear Modeling: Applications and Data Analysis Methods.* Sage Publishing, Newbury Park.

Cook N.R., Rosner B.A., Wei C., Srinivasan S.R., and Berenson G.S. (2004). Using the area under the curve to reduce measurement error in predicting young adult blood pressure from childhood measures. *Statistics in Medicine* **23**, 3421-3435.

Cudeck, R. and Harring, J.R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology* **58**, 615-637.

Cudeck, R. and Klebe, K.J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods* **7**, 41-63.

Davidian M., and Giltinan D.M. (2003). Nonlinear Models for Repeated Measurement Data: An Overview and Update. *Journal of Agricultural, Biological, and Environmental Statistics* **8**, 387-419.

Dietz K. and Hadeler K.P. (1988). Epidemiological models for sexually transmitted diseases. *Journal of Mathematical Biology* **26**, 1-25.

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

Dimidjian, S., Hollon, S.D., Dobson, K.S., Schmaling, K.B., Kohlenberg, R.J., Addis, M.E., Gallop, R., McGlinchey, J.B., Markley, D.K., Gollan, J.K., Atkins, D.C., Dunner, D.L., and Jacobson, N.S. (2006). Behavioral Activation, cognitive therapy, and anti-depressant medication in the acute treatment of major depression. *Journal of Consulting and Clinical Psychology* **74**, 658-370.

Dominicus, A., Ripatti, S., Pederson, N.L., and Palmgren, J. (2006). Modelling variability in longitudinal data using random change point models. *Mathematical Statistics Stockholm University Research Report* **1**, 1-20.

Dominicus, A., Ripatti, S., Pederson, N.L., and Palmgren, J. (2008). A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine* **27**, 5786-5798.

Ertel J.E., and Fowlkes E.B. (1976). Some algorithms for linear spline and piecewise multiple linear regression. *Journal of the American Statistical Association* **71**, 640-648.

Everitt, B.S. (1995). The Analysis of Repeated Measures: A Practical Review with Examples, *The Statistician* **44**, 113-135.

Feder, P.I. (1975). The log likelihood ratio in segmented regression. *The Annals of Statistics* **3**, 84-97.

Gibbons, R.D., Hedeker, D., Elkin, I., Waternaux, C., Kramer, H.C., Greenhouse, J.B., Shea, 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.

Goldstein, H. (1987). *Models in Educational and Social Research.* Oxford University Press, New York.

Hawkins, D.M. (1976). Point estimation of the parameters of piecewise regression models. *Applied Statistics* **25**, 51-57.

Heagerty P.J., and Zeger S.L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science* **75**, 1-26.

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.

Keller, M.B., McCullough, J.P., Klein, D.N. Arnow, B, Dunner, D.L, Gelenberg, A.J., Markowitz, J.C., Nemeroff, C.B., Russell, J.M., Thase, M.E., Trivedi, M.H., and Zajecka, J. (2000). A Comparison of Nefazodone, the Cognitive Behavioral-Analysis system of Psychotherapy, and their combination for the treatment of Chronic Depression. *The New England Journal of Medicine* **342**, 1462-1470.

Kim H.J., Fay M.P., Yu B.B., Barrett M.J., and Feuer E.J. (2004). Comparability of segmented line regression models. *Biometrics* **60**, 1005-1014.

Kuchenhoff H., and Ulm K. (1997). Comparison of statistical methods for assessing threshold limiting values in occupational epidemiology. *Computational Statistics* **12**, 249-264.

Laird, N.M and Ware, J.J. (1982). Random-effect models for longitudinal data. *Biometrics* **38**, 963-974.

Lee, K.J. and Thompson, S.J. (2008). Flexible parametric models for random-effects distributions, *Statistics in Medicine* **27**, 418-434.

Littell R.C., Milliken G.A., Stroup W.W., and Wolfinger R.D. (1995). *SAS System for Mixed Models.* SAS Institute Inc, Cary.

Little R. (1994). Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **88**, 125-34.

Matthews J.N., Altman, D.G., Campbell, M.J., and Royston, P. (1990). Analysis of serial measurements in medical research. *British Medical Journal* **300**, 230-235.

Morrell, C.H., Pearson, J.D., Carter, H.B., and Brant, L.J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association* **90**, 45-53.

Muggeo, V.M.R. (2003). Estimating regression models with unknown breakpoints, *Statistics in Medicine* **22**, 3055-3071.

Muggeo, V.M.R., Atkins, D.C., and Gallop, R. (under review). Segmented mixed models with random changepoints: a maximum likelihood approach with application to depression study in psychology. Manuscript submitted for publication.

Nguyen, H. and Amaratunga, D. (2001). Analysis of Pharmacokinetic Data. In S.P. Millard and A. Krause, editors, *Applied Statistics in the Pharmaceutical Industry.* Springer-Verlag, New York.

Piepho, H.P., and Ogutu, J.O. (2003). Inference for the Break Point in Segmented Regression with Application to Longitudinal Data. *Biometrical Journal* **45**, 591-601.

Powell, M.J.D. (1970). A Survey of Numerical Methods for Unconstrained Optimization. *SIAM Review* **12**, 79-97.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models*, 2nd ed. Sage Publishing, London.

Rigby R.A., and Stasinopoulos D.M. (1992). Detecting break Points in the Hazard Function in Survival Analysis. In *Statistical Modelling*, eds B. Francis, G.U.H. Seeber, P.G.M. van der Heyden and W. Jansen, 303-312. Elsevier Science Publishers B.V.

Robinson, G.K. (1991). That BLUP is a good thing. *Statistical Science*. **6**, 15-51.

Scott, M. A., Norman, R.G., and Berger, K. (2004). Modelling growth and decline in lung function in duchenne's muscolar dystrophy with an augmented linear mixed effects model. *Applied Statistics* **53**, 507-521.

Seber G.A.F., and Wild C.J. (1989). *Nonlinear Regression*. Wiley, New York.

Spritzler, J.; DeGruttola, V.G., and Pei, L. (2008). Two-Sample Tests of Area-Under-the-Curve in the Presence of Missing Data. *The International Journal of Biostatistics* **4**, Issue 1, Article 1.

Stasinopoulos D.M., and Rigby R.A. (1992). Detecting break points in generalised linear models. *Computational Statistics and Data Analysis* **13**, 461-471.

Sternio, J.F., Weisberg, H.I., and Bryk, A.S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics* **39**, 71-86.

Ulm K. (1991). A statistical methods for assessing a threshold in epidemiological studies. *Statistics in Medicine* **10**, 341-349.

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Spinger-Verlag, New York.

Wang, Y. (2007). Derivation of various NONMEM estimation methods. *Journal of Pharmacokinetics and Pharmacodynamics* **34**, 575-593.

Receied October 24, 2009; accepted January 17, 2010.

Robert J. Gallop
West Chester University
Department of Mathematics, Applied Statistics Program,
West Chester, PA 19383
rgallop@wcupa.edu

Sona Dimidjian
University of Colorado at Boulder
Department of Psychology
Boulder, CO 80309
Sona.dimidjian@colordo.edu

David C. Atkins
University of Washington
Department of Psychiatry and Behavioral Science
Seattle, WA 98105
datkins@u.washington.edu

Vito Muggeo
University of Palermo
Dipartimento Scienze Statistiche e Matematiche
Palermo, Italy
Vito.muggeo@unipa.it