

General Approaches to Analysis of Course:  
Applying Growth Mixture Modeling to Randomized  
Trials of Depression Medication

Bengt Muthén  
University of California, Los Angeles

Hendricks C. Brown  
University of Miami

Aimee M. Hunter  
Ian A. Cook  
Andrew F. Leuchter  
University of California, Los Angeles

\*

September 2, 2009.

Forthcoming in P. E. Shrout (ed.), *Causality and Psychopathology:  
Finding the Determinants of Disorders and their Cures*. Washington,  
DC: American Psychiatric Publishing.

---

\*The research of the first author was supported under grant K02 AA 00230-01 from NIAAA, by grant

1 R21 AA10948-01A1 from NIAAA, by grant No. MH40859 from NIMH, and by grant P30 MH066247

## Abstract

This chapter discusses the use of growth mixture modeling to assess treatment effects in clinical trials. The motivation is a study of depression medication in a double-blind placebo-controlled trial. Studies of this type typically show placebo response and placebo non response. Growth mixture modeling (GMM) is well suited for representing such heterogeneity among subjects in that it can identify different types of trajectory shapes. GMM can be seen as a combination of conventional mixed effects modeling and cluster analysis, also allowing prediction of class membership and estimation of each individual's most likely class membership. GMM has particularly strong potential for analyses of randomized trials because it responds to the need to investigate for whom a treatment is effective by allowing for different treatment effects in different trajectory classes. In this trial, a separate analysis of the placebo group finds evidence of a placebo response trajectory class with a strong initial improvement, followed by a later worsening. A separate analysis of the medication group shows two types of responder classes, one with an initial improvement only and one with a sustained improvement. A joint analysis of the placebo and medication groups makes it possible to estimate medication effects in the presence of placebo-response effects and shows benefits of medication. Analysis strategies and alternatives for assessing medication effects are discussed.

Key words: Randomized trials, growth modeling, causal effects, latent variables, trajectory classes, maximum likelihood.

---

from NIDA and NIMH. The email address for the first author is [bmuthen@ucla.edu](mailto:bmuthen@ucla.edu). We thank Tihomir Asparouhov for helpful comments.

# 1 Introduction

This chapter discusses the assessment of treatment effects in longitudinal randomized trials using growth mixture modeling (Muthén & Shedden, 1999; Muthén, & Muthén, 2000; Muthén et al., 2002; Muthén & Asparouhov, 2008). Growth mixture modeling (GMM) is a generalization of conventional repeated measurement mixed effects (multilevel) modeling. GMM captures unobserved subject heterogeneity in trajectories not only by random effects but also by latent classes corresponding to qualitatively different types of trajectories. GMM can be seen as a combination of conventional mixed effects modeling and cluster analysis, also allowing prediction of class membership and estimation of each individual's most likely class membership. GMM has particularly strong potential for analyses of randomized trials because it responds to the need to investigate for whom a treatment is effective by allowing for different treatment effects in different trajectory classes.

The chapter is motivated by a UCLA study of depression medication (Leuchter et al., 2002). Data on 94 subjects are drawn from a combination of three studies carried out with the same design, using three different types of medications: fluoxetine ( $n = 14$ ), venlafaxine IR ( $n = 17$ ), and venlafaxine XR ( $n = 18$ ). Subjects were measured at baseline and again after a 1-week placebo lead-in phase. In the subsequent double-blind phase of the study, the subjects were randomized into a medication ( $n = 49$ ) or placebo group ( $n = 45$ ). After randomization, subjects were measured at nine occasions: at 48 hours and at weeks 1 - 8. The current analyses consider the Hamilton Depression Rating

Scale. Several predictors of course of the Hamilton scale trajectory are available, including gender, treatment history, and a baseline measure of central cordance hypothesized to influence tendency to respond to treatment.

The results of studies of this kind are often characterized in terms of an end-point analysis where the outcome at the end of the study, here at 8 weeks, is considered for the placebo group and for the medication group. A subject may be classified as a responder by showing a week 8 depression score below 10, or when dropping below 50% of the initial score. The treatment effect may be assessed by comparing the medication and placebo group with respect to the ratio of responders to non-responders.

As an alternative to end-point analysis, conventional repeated measurement mixed effects (multilevel) modeling can be used. Instead of focusing on only the last time point, this uses the outcome at all time points, the two pre-treatment occasions and the nine post-treatment occasions. The trajectory shape over time is of key interest and is estimated by a model that draws on the information from all time points. The idea of considering trajectory shape in research on depression medication has been proposed by Quitkin et al. (1991), although not using a formal statistical growth model.

Rates of response to treatment with anti-depressant drugs are estimated as 50 – 60% in typical patient populations. Of particular interest in this chapter is how to assess treatment effects in the presence of placebo response. Placebo response is an improvement in depression ratings that is seen in the placebo group, i.e. unrelated to medication. The improvement is often seen as an early steep drop in depression,

often followed by a later upswing. An example is seen in Figure 2 below. Placebo response confounds the estimation of the true effect of medication and is an important phenomenon given its high prevalence of 25 – 60% (Quitkin, 1999). Because placebo response is pervasive, the statistical modeling must take it into account when estimating medication effects. This can be done by acknowledging the qualitative heterogeneity in trajectory shapes for responders and non responders.

It is important to distinguish among responder and non-responder trajectory shapes in both the placebo and medication groups. Conventional repeated measures modeling may lead to distorted assessment of medication effects when individuals follow several different trajectory shapes. GMM to be considered in this chapter avoids this problem while maintaining the repeated measures modeling advantages. The chapter begins by considering GMM with two classes, a non-responder class and a responder class. The responder class is defined as those individuals who respond in the placebo group and who would have responded to placebo among those in the medication group. The responder class membership is observed for subjects in the placebo group but is unobserved in the medication group. Because of randomization, it can be assumed that this class of subjects is present in both the placebo and medication group and in equal numbers. GMM can identify the placebo responder class in the medication group. Having identified the placebo responder class and placebo non-responder class in both the placebo and medication group, medication effects can more clearly be identified. In one approach, the medication effect is formulated in terms of an effect of medication on the trajectory slopes after the treatment phase has begun. This medication effect is allowed to be different for

the non-responder and responder trajectory classes. Another approach formulates the medication effect as increasing the probability of membership in advantageous trajectory classes and decreasing the probability of membership in disadvantageous trajectory classes.

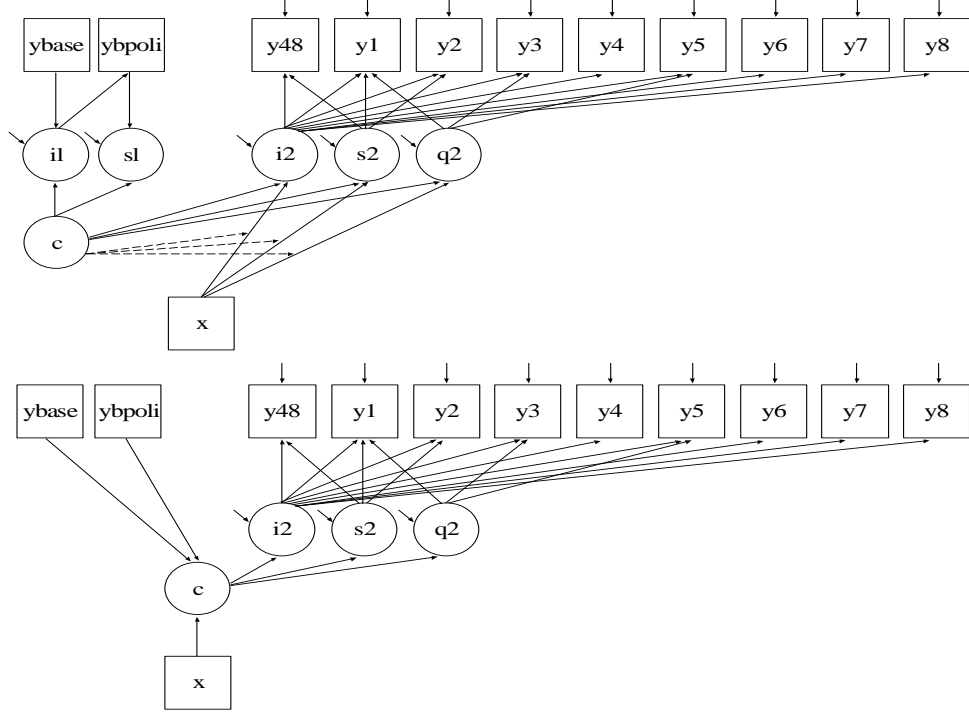
Section 2 describes the statistical model. Section 3 presents analysis results. Section 4 concludes.

## 2 Growth Mixture Modeling

This section gives a brief description of the growth mixture model (GMM) in the context of the current study. A two-piece, random effect growth mixture model is applied to the Hamilton Depression Rating Scale outcomes at the 11 time points,  $y_1 - y_{11}$ . The first piece refers to the two time points  $y_1, y_2$  before randomization and the second piece refers to the nine post-randomization time points  $y_3 - y_{11}$ . Given only two time points, the first piece is by necessity taken as a linear model with a random intercept, defined at baseline, and a fixed effect slope. An exploration of each individual's trajectory suggests a quadratic trajectory shape for the second piece. The growth model for the second piece is centered at week 8, defining the random intercept as the systematic variation at that time point. All random effect means are specified as varying across latent trajectory classes. The medication effect is captured by a regression of the linear and quadratic slopes in the second piece on a medication dummy variable. These medication effects are allowed to vary across the latent trajectory classes. The model is shown in diagrammatic

Figure 1: Two alternative GMM approaches

## Figure 1. Two alternative GMM approaches



form at the top of Figure 1.<sup>1</sup>.

The statistical specification is as follows. Consider the depression outcome  $y_{it}$  for individual  $i$ , let  $c$  denote the latent trajectory class variable, let  $\eta$  denote random effects, let  $a_t$  denote time, and let  $\epsilon_t$  denote residuals containing measurement error and time-specific variation. For the first, pre-randomization piece, conditional on trajectory class

---

<sup>1</sup>In Figure 1 the observed outcomes are shown in boxes and the random effects in circles. Here,  $i$ ,  $s$ , and  $q$  denote intercept, linear slope, and quadratic slope, respectively. In the formulas below, these random effects are referred to as  $\eta_0$ ,  $\eta_1$ , and  $\eta_2$ . The treatment dummy variable is denoted  $x$ .

$k$  ( $k = 1, 2, \dots, K$ ),

$$y_{it}^{pre}|_{c_i=k} = \eta_{0i}^{pre} + \eta_{1i}^{pre} a_t + \epsilon_{it}^{pre}, \quad (1)$$

with  $a_1 = 0$  to center at baseline, and random effects

$$\eta_{10i}^{pre}|_{c_i=k} = \alpha_{10k} + \zeta_{10i}, \quad (2)$$

$$\eta_{11i}^{pre}|_{c_i=k} = \alpha_{11k} + \zeta_{11i}. \quad (3)$$

With only two pre-randomization time points, the model is simplified to assume a non-random slope,  $V(\zeta_{11}) = 0$ , for identification purposes. For the second, post-randomization piece,

$$y_{it}|_{c_i=k} = \eta_{0i} + \eta_{1i} a_t + \eta_{2i} a_t^2 + \epsilon_{it}. \quad (4)$$

Here,  $a_{11} = 0$  defining  $\eta_{0i}$  as the week 8 depression status. The remaining  $a_t$  values are set according to the distance in timing of measurements. Assume for simplicity a single drug and denote the medication status for individual  $i$  by the dummy variable  $x_i$  ( $x = 0$  for the placebo group and  $x = 1$  for the medication group).<sup>2</sup> The random effects are allowed to be influenced by group and a covariate  $w$ , their distributions varying as a function of trajectory class ( $k$ ),

$$\eta_{0i}|_{c_i=k} = \alpha_{0k} + \gamma_{01k} x_i + \gamma_{02k} w_i + \zeta_{0i}, \quad (5)$$

$$\eta_{1i}|_{c_i=k} = \alpha_{1k} + \gamma_{11k} x_i + \gamma_{12k} w_i + \zeta_{1i}, \quad (6)$$

$$\eta_{2i}|_{c_i=k} = \alpha_{2k} + \gamma_{21k} x_i + \gamma_{22k} w_i + \zeta_{2i}. \quad (7)$$

The residuals  $\zeta_i$  in the first and second piece have a  $4 \times 4$  covariance matrix  $\Psi_k$ , here taken to be constant across classes  $k$ . For both pieces the residuals  $\epsilon_{it}$  have a  $T \times T$

---

<sup>2</sup>In the application three dummy variables are used to represent the three different medications.



covariance matrix  $\Theta_k$ , here taken to be constant across classes. For simplicity,  $\Psi_k$  and  $\Theta_k$  are assumed to not vary across treatment groups. As seen in (5) - (7), the placebo group ( $x_i = 0$ ) consists of subjects from the two different trajectory classes that vary in the means of the growth factors, which in the absence of a covariate  $w$  are represented by  $\alpha_{0k}$ ,  $\alpha_{1k}$ , and  $\alpha_{2k}$ . This gives the average depression development in the absence of medication. Because of randomization, the placebo and medication group are assumed to be statistically equivalent at the first two time points. This implies that  $x$  is assumed to have no effect on  $\eta_{10i}$  or  $\eta_{11i}$  in the first piece of the development. Medication effects are described in the second piece by  $\gamma_{01k}$ ,  $\gamma_{11k}$ , and  $\gamma_{21k}$  as a change in average growth rate that can be different for the classes.

This model allows the assessment of medication effects in the presence of placebo response. A key parameter is the medication-added mean of the intercept random effect centered at week 8. This is the  $\gamma_{01k}$  parameter of (5). This indicates how much lower or higher the average score is at week 8 for the medication group relative to the placebo group in the trajectory class considered. In this way, the medication effect is specific to classes of individuals who would or would not have responded to placebo. The modeling will be extended to allow for the three drugs of this study to have different  $\gamma$  parameters in (5) - (7).

Class membership can be influenced by baseline covariates as expressed by a logistic regression, e.g. with two classes,

$$\log[P(c_i = 1|x_i)/P(c_i = 2|x_i)] = \alpha_c + \gamma_c w_i, \quad (8)$$

where  $c = 1$  may refer to the non-responder class and  $c = 2$  the responder class. It may be noted that this model assumes that medication status does not influence class membership. Class membership is conceptualized as a quality characterizing an individual before entering the trial.

A variation of the modeling will focus on post-randomization time points. Here, an alternative conceptualization of class membership is used. Class membership is thought of as being influenced by medication so that the class probabilities are different for the placebo group and the three medication groups. Here, the medication effect is quantified in terms of differences across groups in class probabilities. This model is shown in diagrammatic form at the bottom of Figure 1. It is seen that the GMM involves only the post-randomization outcomes which is logical given that treatment influences the latent class variable which in turn influences the post-treatment outcomes. In addition to the treatment variable, pre-treatment outcomes may be used as predictors of latent class as indicated in the figure. The treatment and pre-treatment outcomes may interact in their influence on latent class membership.

## 2.1 Estimation and Model Choice

The GMM can be fitted into the general latent variable framework of the Mplus program (Muthén & Muthén, 1998 - 2008). Estimation is carried out using maximum-likelihood via an EM algorithm. Missing data under the MAR assumption are allowed for the outcomes. Given an estimated model, estimated posterior probabilities for each indi-

vidual and each class are produced. Individuals can be classified into the class with highest probability. The classification quality is summarized in an entropy value with range  $0 - 1$ , where 1 corresponds to the case where all individuals have probability 1 for one class and 0 for the others. For model fitting strategies, see Muthén et al. (2002), Muthén (2004), and Muthén and Asparouhov (2008). A common approach to decide on the number of classes is to use the Bayesian information criterion (BIC) which puts a premium on models with large loglikelihood values and small number of parameters. The lower the BIC, the better the model. Analysis of depression trial data have an extra difficulty due to the typically small sample sizes. Little is known about the performance of BIC for samples as small as in the current study. Bootstrapped likelihood ratio testing can be performed in Mplus (Muthén & Asparouhov, 2008), but the power of such testing may not be sufficient at these sample sizes. Plots showing the agreement between the class-specific estimated means and the individual trajectories for individuals most likely belonging to a class can be useful in visually inspecting models, but are only of limited value in choosing between models.

A complication of maximum-likelihood GMM is the presence of local maxima. These are more prevalent with smaller samples such as the current ones for the placebo group, the medication group, as well as for the combined sample. To be confident that a global maximum has been found, many random starting values need to be used and the best loglikelihood value replicated several times. In the present analyses, between 500 and 4000 random starts were used depending on the complexity of the model.

### 3 Growth Mixture Analyses

In this section the depression data are analyzed in three steps using GMM. First, the placebo group is analyzed alone. Second, the medication group is analyzed alone. Third, the placebo and medication groups are analyzed jointly according to the growth mixture model just presented in order to assess the medication effects.

#### 3.1 Analysis of the Placebo Group

A 2-class GMM analysis of the 45 subjects in the placebo group resulted in the model-estimated mean curves shown in Figure 2. As expected, a responder class (class 1) shows a post-randomization drop in the depression score with a low of 7.9 at week 5 and with an upswing to 10.8 at week 8. An estimated 32% of the subjects belong to the responder class. In contrast, the non-responder class has a relatively stable level for weeks 1- 8, ending with a depression score of 15.6 at week 8. The sample standard deviation at week 8 is 7.6. It may be noted that the baseline score is only slightly higher for the non-responder class, 22.7 versus 21.9. The standard deviation at baseline is 3.6.<sup>3</sup>

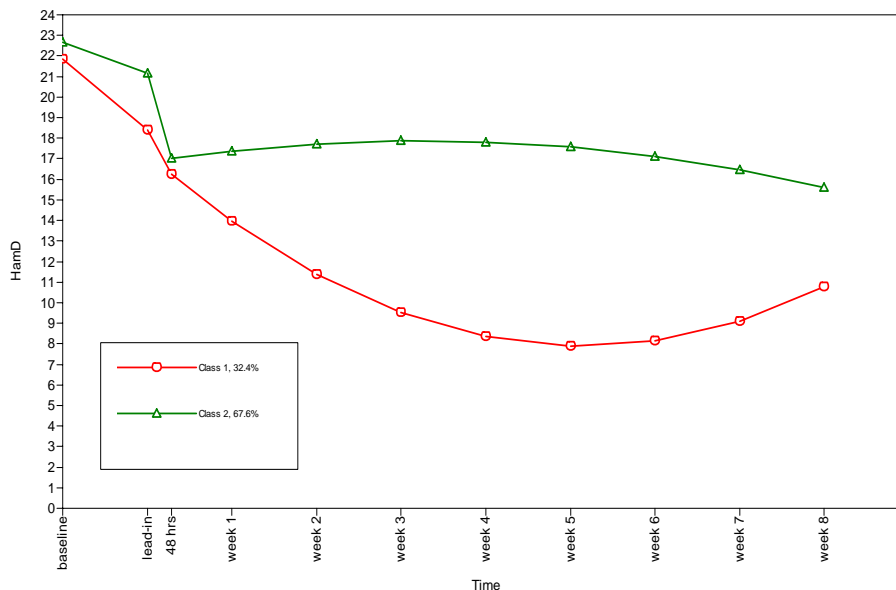
The observed trajectories of individuals classified into the two classes are plotted in Figure 3a and b as broken lines, whereas the solid curves show the model-estimated means. The figure indicates that the estimated mean curves represent the individual

---

<sup>3</sup>The maximum loglikelihood value for the 2-class GMM of Figure 2 is 1,055.974, which is replicated across many random starts, with 28 parameters and a BIC value of 2,219. The classification based on the posterior class probabilities is not clearcut in that the classification entropy value is only 0.66.

Figure 2: Two-class GMM for placebo group

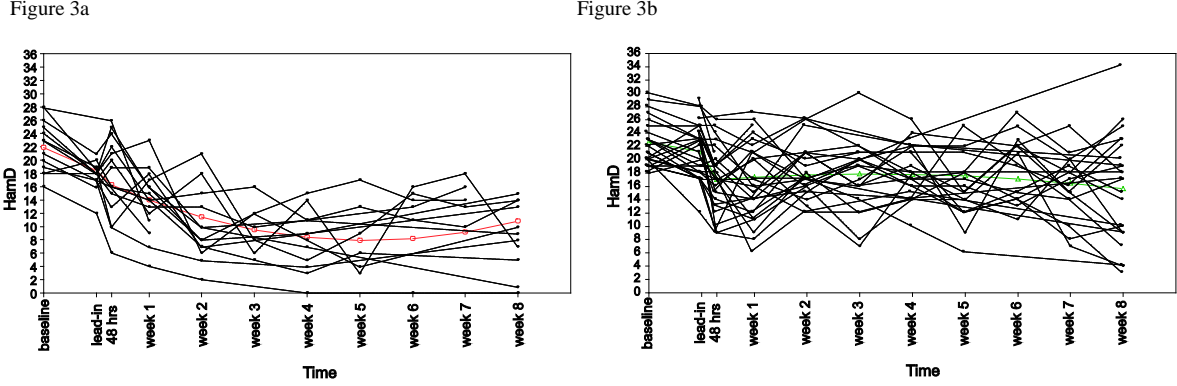
Figure 2



development rather well, although there is a good amount of individual variation around the mean curves.

It should be noted that the classification of subjects based on the trajectory shape approach of GMM will not agree with that of using end-point analysis. As an example, the non-responder class of Figure 3b shows two subjects with scores less than 5 at week 8. The individual with the lowest score at week 8, however, has a trajectory that agrees well with the non-responder mean curve for most of the trial, only deviating from it during the last two weeks. The week 8 score has a higher standard deviation than at earlier time points, thereby weighting this time point somewhat less. Also, the data coverage due to missing observations is considerably lower for weeks 5 - 7 than other weeks, reducing the weight of these time points. The individual with the second-lowest

Figure 3: Individual trajectories for placebo subjects classified into (a) the responder class and (b) the non-responder class



score at week 8 deviates from the mean curve for week 5 but has missing data for weeks 6 and 7. This person is also ambiguously classified in terms of his/her posterior probability of class membership.

To further explore the data, a 3-class GMM was also fitted to the 45 placebo subjects. Figure 4a shows the mean curves for this solution. This solution no longer shows a clear-cut responder class. Class 2 (49%) declines early, but the mean score does not go below 14. Class 1 (22%) ends with a mean score of 10.7, but does not show the expected responder trajectory shape of an early decline.<sup>4</sup> A further analysis was made to investigate if the lack of a clear responder class in the 3-class solution is due to the sample size of  $n = 45$  being too small to support three classes. In this analysis, the  $n = 45$  placebo group subjects were augmented by the medication group subjects but using only the two pre-randomization time points from the medication group. Because

<sup>4</sup>The loglikelihood value for the model in Figure 4a is 1,048,403, replicated across several random starts, with 34 parameters and a BIC value of 2,226. Although the BIC value is slightly worse than for the 2-class solution, the classification is better as shown by the entropy value of 0.85.

Figure 4: (a) Three-class GMM for placebo group. (b) Three-class GMM for placebo

group and pre-randomization medication group individuals

Figure 4a

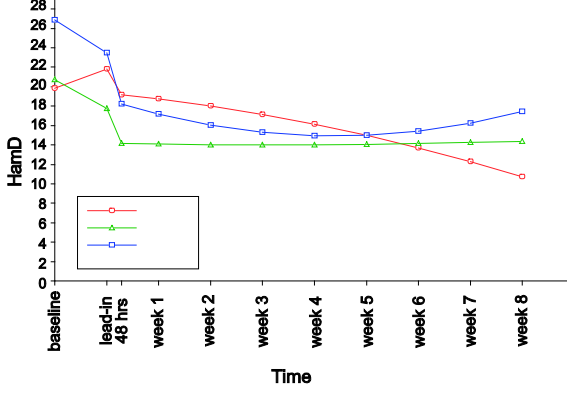
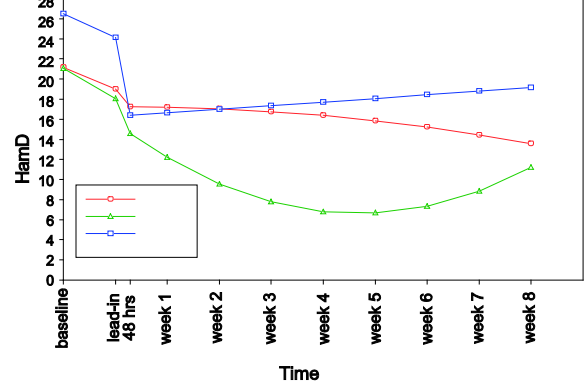


Figure 4b



of randomization, subjects are statistically equivalent before randomization, so this approach is valid. The first, pre-randomization piece of the GMM has nine parameters, leaving only 25 parameters to be estimated in the second, post-randomization piece by the  $n = 45$  placebo subjects alone. Figure 4b shows that a responder class (class 2) is now found with 21% of the subjects estimated to be in this class. A high (class 3) and a low (class 1) non-responder class is found with 18% and 60% estimated to be in these classes, respectively. As compared to Figure 3, the observed individual trajectories within class are somewhat less heterogeneous (trajectories not shown).<sup>5</sup>

<sup>5</sup>The loglikelihood value for the model in Figure 4b is 1,270.030, replicated across several random starts, with 34 parameters and a BIC value of 2,695. The entropy value is 0.62. Because a different sample size is used, these values are not comparable to the earlier ones.

## 3.2 Analysis of the Medication Group

Two major types of GMMs were applied to the medication group. A first type analyzes all time points and either makes no distinction among the three drugs (fluoxetine, venlafaxine IR, venlafaxine XR), or allows drug differences for the class-specific random effect means of the second piece of the GMM. It would not make sense to also let class membership vary as a function of drug since class membership is conceptualized as a quality characterizing an individual before entering the trial. Class membership influences pre-randomization outcomes, which cannot be influenced by drugs.

To investigate class membership, a second type of GMM analyzes the nine post-randomization time points to both focus on the period where the medications have an effect and to let the class membership correspond to only post-randomization variables. Here, differences across the three drugs are allowed not only for the random effect means for each of the classes, but the drug type is also allowed to influence class probabilities.

### 3.2.1 Analysis of All Time Points

A 2-class GMM analysis of the 49 subjects in the medication group resulted in the model-estimated mean curves shown in Figure 5. As expected, one of the classes is a large responder class (class 1, 85%). The other class (class 2, 15%) improves initially, but then worsens.<sup>6</sup>

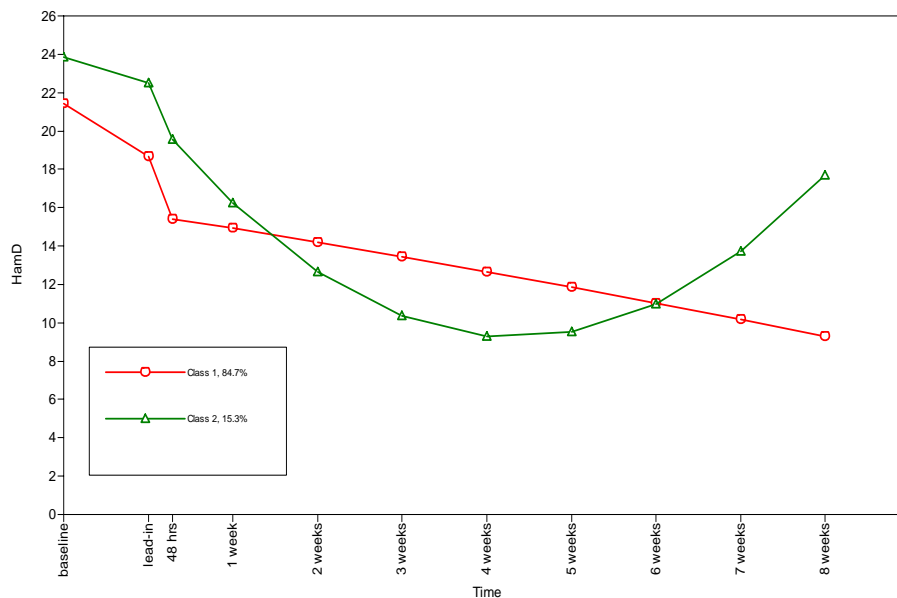
---

<sup>6</sup>The loglikelihood value for the model in Figure 5 is  $-1,084.635$ , replicated across many random starts, with 28 parameters and a BIC value of 2,278. The entropy value is 0.90.



Figure 5: Two-class GMM for medication group

Figure 5



A 3-class GMM analysis of the 49 subjects in the medication group resulted in the model-estimated mean curves shown in Figure 6. The three mean curves show the expected responder class (class 3, 68%) and the class (class 2, 15%) found in the 2-class solution showing an initial improvement, but later worsening. In addition, a non-response class (class 1, 17%) emerges which has no medication effect throughout.<sup>7</sup>

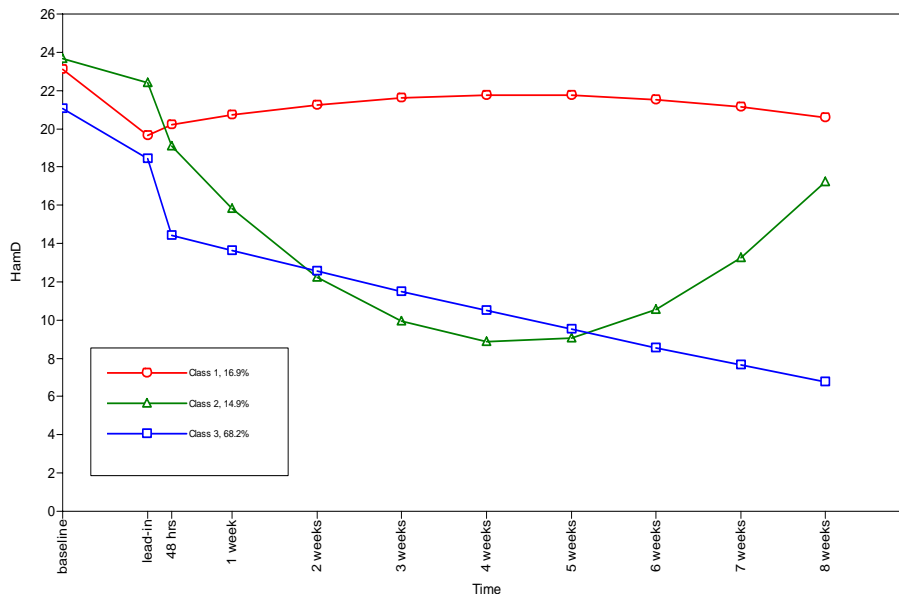
Allowing for drug differences for the class-specific random effect means of the second piece of the GMM did not give a trustworthy solution in that the best loglikelihood value was not replicated. This may be due to the fact that this model has more parameters

---

<sup>7</sup>The loglikelihood value for the model in Figure 6 is  $-1,077.433$ , replicated across many random starts, with 34 parameters and a BIC value of 2,287. The BIC value is worse than for the 2-class solution. The entropy value is 0.85.

Figure 6: Three-class GMM for medication group

Figure 6



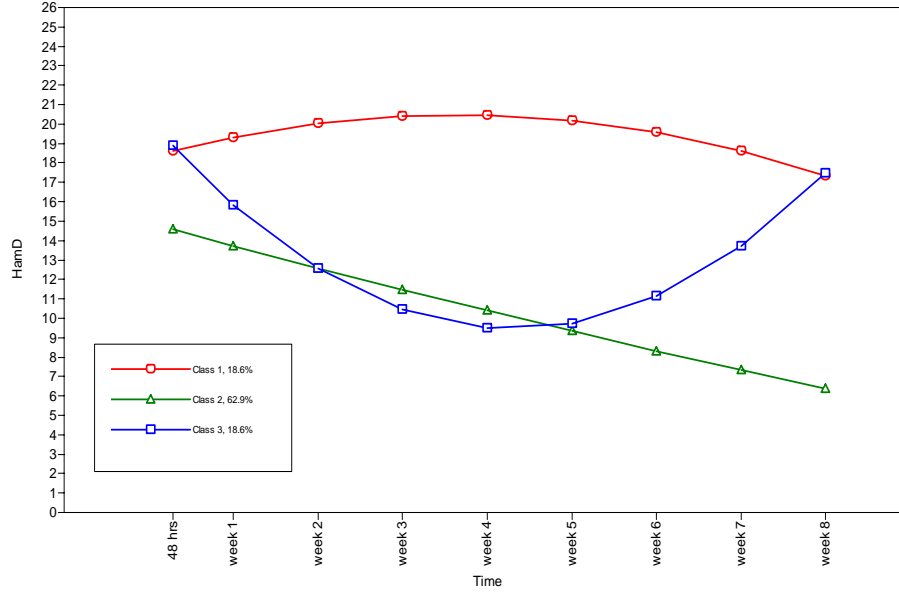
than subjects (59 versus 49).

### 3.2.2 Analysis of Post-Randomization Time Points

As a first step, 2- and 3-class analysis of the nine post-randomization time points were performed not allowing for differences across the three drugs. This gave solutions that were very similar to those of Figure 5 and Figure 6. The similarity in mean trajectory shape held up also when allowing for class probabilities to vary as a function of drug. Figure 7 shows the estimated mean curves for this latter model. The estimated class probabilities for the three drugs show that in the responder class (class 2, 63%) 21% of the subjects are on fluoxetine, 29% are on venlafaxine IR, and 50% are on venlafaxine XR. For the non-responder class that shows an initial improvement and a later wors-

Figure 7: Three-class GMM for medication group post randomization

Figure 7



ening (class 3, 19%), 25% are on fluoxetine, 75% are on venlafaxine IR, and 0% are on venlafaxine XR. For the non-responder class that shows no improvement at any point (class 1, 19%), 58% are on fluoxetine, 13% are on venlafaxine IR, and 29% are on venlafaxine XR. Judged across all three trajectory classes, this suggests that in venlafaxine XR has the better outcome, followed by venlafaxine IR, and with fluoxetine last. Note, however, that for these data subjects were not randomized to the different medications and therefore comparisons among medications are confounded by subject differences.<sup>8</sup>

As a second step, a 3-class model was analyzed by a GMM, where not only class membership probability was allowed to vary across the three drugs, but also the class-

---

<sup>8</sup>The loglikelihood value for the model of Figure 7 is  $-873.831$ , replicated across many random starts, with 27 parameters and a BIC value of 1,853. The entropy value is 0.79.

Figure 8: Three-class GMM for (a) fluoxetine subjects, (b) venlafaxine IR subjects, and (c) venlafaxine XR subjects

Figure 8a

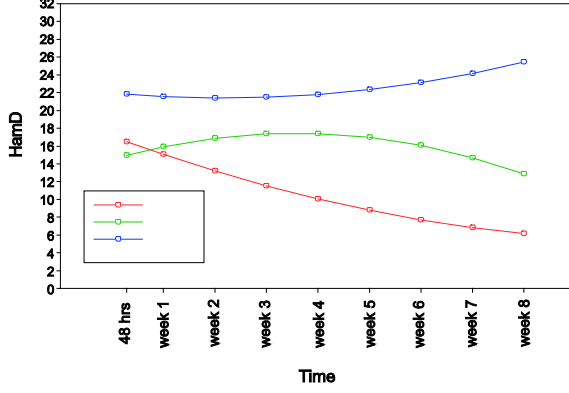


Figure 8b

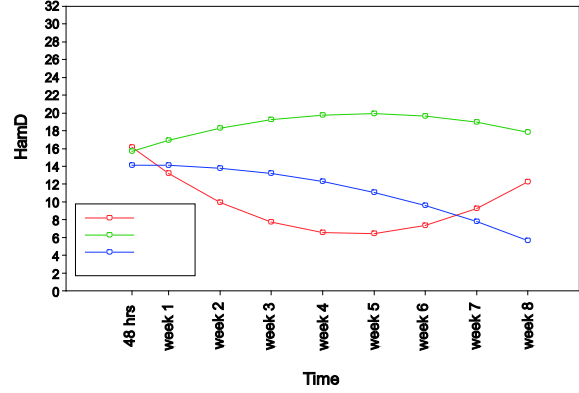
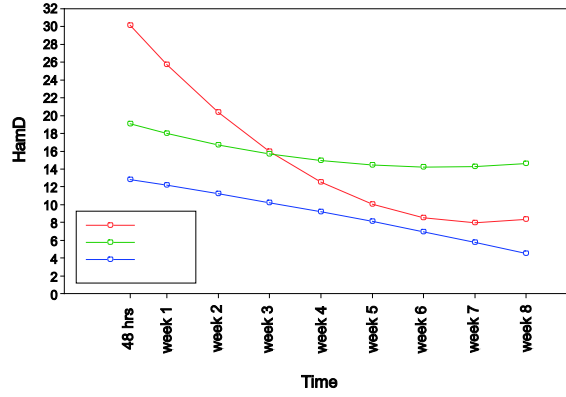


Figure 8c



varying random effect means. This analysis showed no significant drug differences in class membership probabilities. As shown in Figure 8 the classes are essentially of different nature for the three drugs.<sup>9</sup>

<sup>9</sup>The loglikelihood value for the model of Figure 8 is  $-859.577$ , replicated in only a few random starts, with 45 parameters and a BIC value of 1,894. The entropy value is 0.81. It is difficult to choose between the model of Figure 7 and the model of Figure 8 based on statistical indices. The Figure 7 model has the better BIC value, but the improvement in the loglikelihood of the Figure 8 model is substantial.

### **3.3 Analysis of Medication Effects Taking Placebo Response into Account**

The separate analyses of the 45 subjects in the placebo group and the 49 subjects in the medication group provide the basis for the joint analysis of all 94 subjects. Two types of GMMs will be applied. The first is directly in line with the model shown in Section 2, where medication effects are conceptualized as post-randomization changes in the slope means. The second type uses only the post-randomization time points and class membership is thought of as being influenced by medication in line with the Figure 7 model. Here, the class probabilities are different for the placebo group and the three medication groups so that medication effect is quantified in terms of differences across groups in class probabilities.

#### **3.3.1 Analysis of All Time Points**

For the analysis based on the Section 2 model, a 3-class GMM will be used given that three classes were found interpretable for both the placebo and the medication groups. Figure 9 shows the estimated mean curves for the 3-class solution for the placebo group, the fluoxetine group, the venlafaxine IR group, and the venlafaxine XR group. It is interesting to note that for the placebo group, the Figure 9a mean curves are similar in shape to those of Figure 4b, although the responder class (class 3) is now estimated as 34%. Note that for this model the class percentages are specified to be the same in the medication groups as in the placebo group. The estimated mean curves for the

Figure 9: Three-class GMM of both groups: (a) Placebo subjects, (b) fluoxetine subjects, (c) venlafaxine IR subjects, and (d) venlafaxine XR subjects

Figure 9a

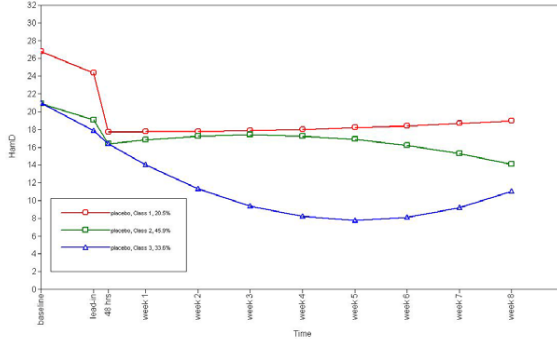


Figure 9b

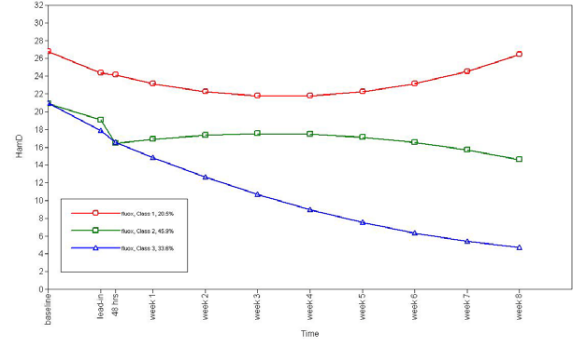


Figure 9c

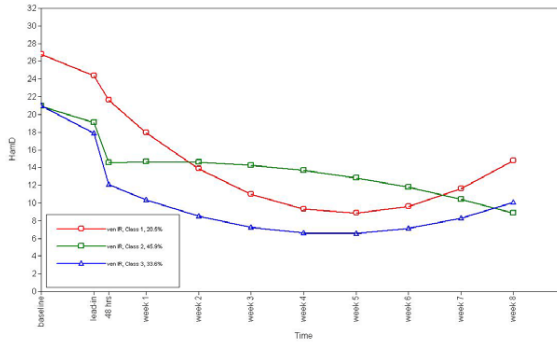
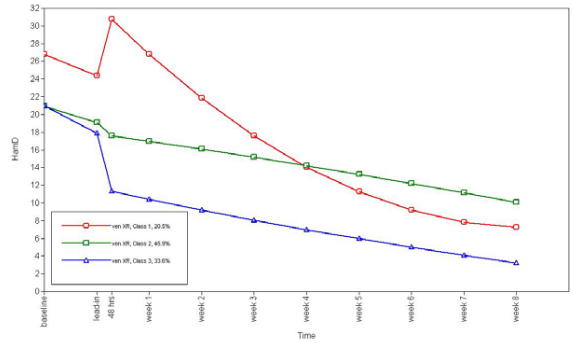


Figure 9d



three medication groups shown in Figure 9b, c, and d are similar in shape to those of the medication group analysis shown in Figure 8a, b, and c. These agreements with the separate-group analyses strengthen the plausibility of the modeling.

This model allows the assessment of medication effects in the presence of placebo response. A key parameter is the medication-added mean of the intercept random effect centered at week 8. This is the  $\gamma_{01k}$  parameter of (5) in Section 2. For a given trajectory class, this indicates how much lower or higher the average score is at week 8 for the medication group in question relative to the placebo group. In this way, the medication

effect is specific to classes of individuals who would or would not have responded to placebo.

The  $\gamma_{01k}$  estimates of the Figure 9 model are as follows. The fluoxetine effect for the high non-responder class 1 at week 8 as estimated by the GMM is significantly positive (higher depression score than for the placebo group), 7.4, indicating a failure of this medication for this class of subjects. In the low non-responder class 2 the fluoxetine effect is small positive and insignificant. In the responder class, the fluoxetine effect is significantly negative (lower depression score than for the placebo group),  $-6.3$ . The venlafaxine IR effect is insignificant for all three classes. The venlafaxine XR effect is significant negative  $-11.7$  for class 1, which after an initial slight worsening turns into a responder class for venlafaxine XR. For the non-responder class 2 the venlafaxine XR effect is insignificant, while for the responder class it is significant negative  $-7.8$ . In line with the medication group analysis shown in Figure 7, the joint analysis of placebo and medication subjects indicates that venlafaxine XR has the most desirable outcome relative to the placebo group. None of the drugs is significantly effective for the low non-responder class 2.<sup>10</sup>

### 3.3.2 Analysis of Post-Randomization Time Points

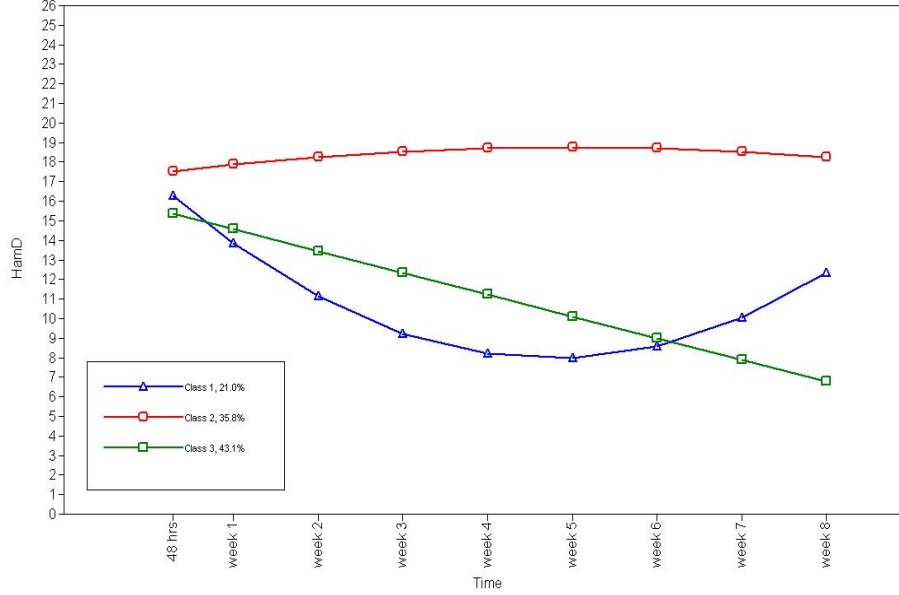
As a final analysis the placebo and medication groups were analyzed together for the post-randomization time points. Figure 10 displays the estimated 3-class solution which

---

<sup>10</sup>The loglikelihood value for the model shown in Figure 9 is  $-2,142.423$ , replicated across a few random starts, with 61 parameters and a BIC value of 4,562. The entropy value is 0.76.

Figure 10: Three-class GMM analysis of both groups using post-randomization time points

Figure 10



again shows a responder class, a non-responder class which initially improves but then worsens (similar to the placebo response class found in the placebo group), and a high non-responder class.<sup>11</sup> As a first step, it is of interest to compare the joint placebo-medication group analysis of Figure 10 to the separate placebo group analysis of Figure 4b and the separate medication group analysis of Figure 6.

Comparing the joint analysis in Figure 10 to that of the placebo group analysis of Figure 4b indicates the improved outcome when medication group individuals are added to the analysis. In the placebo group analysis of Figure 4b 78% are in the two highest,

<sup>11</sup>The loglikelihood value for the model shown in Figure 10 is  $-1,744.999$ , replicated across many random starts, with 29 parameters and a BIC value of 3,621. The entropy value is 0.69.

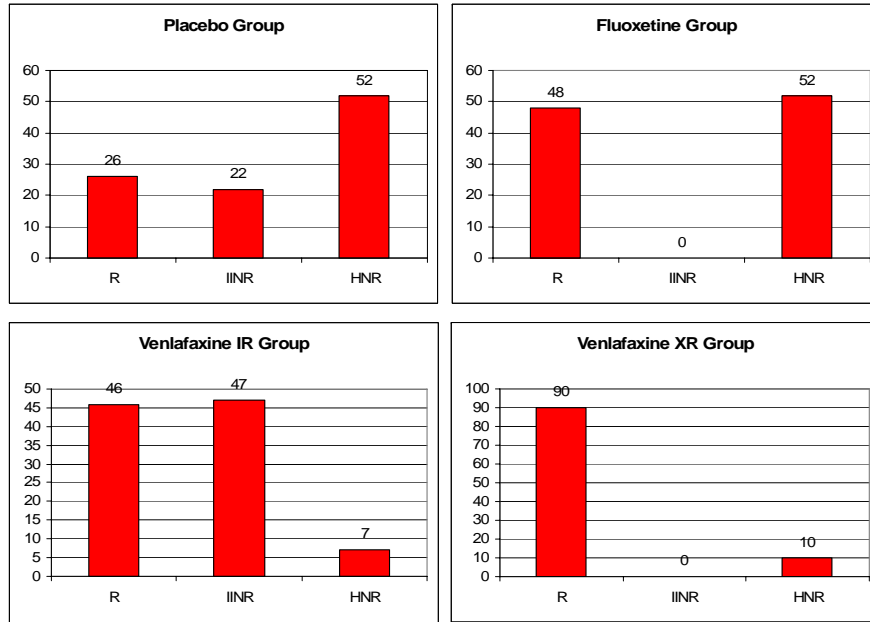


clearly non-responding trajectory classes, whereas in the joint analysis of Figure 10 only 36% are in the highest, clearly non-responding class. In this sense, medication seems to have a positive effect in reducing depression. Furthermore, in the placebo analysis, 21% are in the placebo-responding class which ultimately worsens, whereas in the joint analysis 21% are in this type of class while 43% are in a clearly responding class.

Comparing the joint analysis in Figure 10 to that of the medication group analysis of Figure 6 indicates the worsened outcome when placebo group individuals are added to the analysis. In the medication group analysis of Figure 6 only 17% are in the non-responding class compared to 36% in the joint analysis of Figure 10. Figure 6 shows 15% in the initially improving but ultimately worsening class compared to 21% in Figure 10. Figure 6 shows 68% in the responding class compared to 43% in Figure 10. All three of these comparisons indicate that medication has a positive effect in reducing depression.

As a second step, is of interest to study the medication effects for each medication separately. The joint analysis model allows this because the class probabilities differ between the placebo group and each of the three medication groups as expressed by (8). The results are shown in Figure 11. For the placebo group, the responder class (class 3) is estimated as 26%, the initially improving non-responder class (class 1) as 22%, and the high non-responder class (class 2) as 52%. In comparison, for the fluoxetine group the responder class is estimated as 48% (better than placebo), the initially improving non-responder class as 0% (better than placebo), and the high non-responder class as 52% (same as placebo). For the venlafaxine IR group, the responder class is estimated as 46% (better than placebo), the initially improving non-responder class as 47% (worse

**Figure 11. Medication effects in each of 3 trajectory classes**



R = Responder Class

IINR = Initially Improving Non-Responder Class

HNR = High Non-Responder Class

than placebo), and the high non-responder class as 7% (better than placebo). For the venlafaxine XR group, the responder class is estimated as 90% (better than placebo), the initially improving non-responder class as 0% (better than placebo), and the high non-responder class as 10% (better than placebo).

## 4 Conclusions

The growth mixture analysis presented here demonstrates that, unlike conventional repeated measures analysis, it is possible to estimate medication effects in the presence of placebo effects. The analysis is flexible in that the medication effect is allowed to differ across trajectory classes. This approach should therefore have wide applicability in clinical trials. It was shown that medication effects could be expressed as causal effects. The analysis also produces a classification of individuals into the trajectory classes.

Medication effects were expressed in two alternative ways, as changes in growth slopes and as changes in class probabilities. Related to the latter approach a possible generalization of the model is to include two latent class variables, one before and one after randomization, and letting the medication influence the post-randomization latent class variable as well as transitions between the two latent class variables. Another generalization is proposed in Muthén and Brown (2009) considering four classes of subjects: (1) subjects who would respond to both placebo and medication, (2) subjects who would respond to placebo but not to medication, (3) subjects who would respond to medication but not placebo, and (4) subjects who would respond to neither placebo nor medication. Class (3) is of particular interest from a pharmaceutical point of view.

Prediction of class membership can be incorporated as part of the model, but was not explored here. Such analyses suggest interesting opportunities for designs of trials. If at baseline an individual is predicted to belong to a non-responder class, a different treatment can be chosen.

## References

- Muthén, B. & Brown, C.H. (2009). Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. Forthcoming in *Statistics in Medicine*.
- Leuchter, A. F., Cook, I.A., Witte, E.A., Morgan, M., & Abrams, M. (2002). Changes in brain function of depressed subjects during treatment with placebo. *American Journal of Psychiatry*, 159, 122-129.
- Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L.M. & Sayer, A. (eds.), *New Methods for the Analysis of Change* (pp. 291-322). Washington, D.C.: APA.
- Muthén, B. (2001b). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.
- Muthén B. & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molen-

- berghs, G. (eds.), *Longitudinal Data Analysis*, pp. 143-165. Boca Raton: Chapman & Hall/CRC Press.
- Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3, 459-475.
- Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24, 882-891.
- Muthén, B. & Muthén, L. (1998-2008). *Mplus User's Guide*. Fifth edition. Los Angeles, CA: Muthén & Muthén.
- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.