Bayesian Non-Parametric Hierarchical Modeling for Multiple Membership Data in Grouped Attendance Interventions

Terrance D. Savitsky*and Susan M. Paddock*

December 6, 2012

Abstract

We develop a dependent Dirichlet process (DDP) model for repeated measures multiple membership (MM) data. This data structure arises in studies under which an intervention is delivered to each client through a sequence of elements which overlap with those of other clients on different occasions. Our interest concentrates on study designs for which the overlaps of sequences occur for clients who receive an intervention in a shared or grouped fashion whose memberships may change over multiple treatment events. Our motivating application focuses on evaluation of the effectiveness of a group therapy intervention with treatment delivered through a sequence of cognitive behavioral therapy session blocks, called modules. An open-enrollment protocol permits entry of clients at the beginning of any new module in a manner that may produce unique MM sequences across clients. We begin with a model that composes

*RAND Corporation, 1776 Main Street, Box 2138, Santa Monica, CA 90401-2138 USA

an addition of client and multiple membership module random effect terms, which are assumed independent. Our MM DDP model relaxes the assumption of conditionally independent client and module random effects by specifying a collection of random distributions for the client effect parameters that are indexed by the unique set of module attendances. We demonstrate how this construction facilitates examining heterogeneity in the relative effectiveness of group therapy modules over repeated measurement occasions.

Key words: Bayesian hierarchical models, Conditional autoregressive prior, Dependent Dirichlet process, Group therapy, Growth curve, Mental health, Multiple membership, Non-parametric priors, Substance abuse treatment

1 Introduction

For many applications in which data have a multilevel structure, observations on a study participant might not be nested within a single higher-level unit. Multiple membership (MM) modeling is used to account for such data structures, which arise in applications such as the estimation of teacher effects from student test scores, where each student is typically linked to multiple teachers over one or more grades (Hill & Goldstein 1998). MM structures also occur in the analysis of health care costs when patients are treated by multiple providers (Carey 2000) and smoothing disease rates when modeling health outcomes across geographic areas (Langford et al. 1999).

In our motivating application, the MM structure arises in a study of the effect of group cognitive behavioral therapy (CBT) on reducing depressive symptoms among clients in residential substance abuse treatment. The Building Recovery by Improving Goals, Habits and Thoughts (BRIGHT) study (Watkins et al. 2011) was a community-based effectiveness trial of a group cognitive behavioral therapy (CBT) intervention

for treating residential substance abuse treatment clients having depressive symptoms. The BRIGHT study employed a quasi-experimental design in which cohorts of clients at each of four study sites received either residential treatment as usual (UC) ($n = 159$) or residential treatment enhanced with the BRIGHT intervention (CBT) provided by trained substance abuse treatment counselors ($n = 140$). Clients were assigned to receive either CBT or UC according to which intervention was offered at their study sites at the time of entry into residential substance abuse treatment. CBT and UC were offered at each study site on an alternating basis over time. The clients assigned to the CBT condition were expected to complete four modules of group CBT, with each module consisting of four thematically-similar sessions offered over a two-week period. This sequence of modules was then offered on repeating basis. In all, $S = 61$ group CBT modules were offered to the clients assigned to the CBT condition. These 61 modules were divided into $G = 4$ CBT open-enrollment therapy groups, which are sequences of sessions that have distinct sets of clients; the number of clients enrolled in each open-enrollment group was 17, 21, 19, and 83, respectively. Enrollment into the therapy group occurs on an open basis (Morgan-Lopez & Fals-Stewart 2006, Paddock et al. 2011), with clients entering the therapy group at the start of new modules. The primary study outcome is client depressive symptomology, as measured by the Beck Depression Inventory-II (BDI-II) (Beck et al. 1996). The BDI-II score is a sum across 21 four-level items (scored $0 - 3$), with a higher score indicating a greater level of depressive symptoms. The BDI-II score for client i is measured up to o_i times, with $o_i = 1$ for clients with only a baseline assessment at study entry and up to $o_i = 3$ for clients measured as well at both 3 and 6 months post-baseline. The MM structure arises here since client outcomes might be correlated due to common module attendance, and the BDI-II scores are not uniquely associated with a single module but rather with all modules attended by a client.

For longitudinal studies in which participants belong to multiple higher-level units, the standard analytic approach is to include a single set of random effects terms that are assumed to be constant over time to account for the multiple membership. However, constraining these random effects to be constant across time does not allow for changes in correlations among outcomes for clients who attend modules together; their outcomes might be more strongly correlated immediately following group therapy versus at baseline or longer-term follow-up times. Further, including distinct terms in the model to account for multiple membership and for the correlation of repeated measurements within-client might be too restrictive for applications such as group cognitive behavioral therapy (CBT). Not all clients benefit similarly from group therapy (Smokowski et al. 2001). For example, group climate and cohesion are associated with improved outcomes (Ryum et al. 2009, Crowe & Grenyer 2008). Thus, not only might the effects of modules change over time, but the effects of modules on participant outcome trajectories might vary across study participants.

We present a dependent Dirichlet process (DDP) model for repeated measures multiple membership data. Specifically, we propose a set of random distributions for client random effect parameters that are indexed by therapy group module attendance sequences. Our model allows one to obtain treatment effect estimates for group therapy versus a comparison condition that account for the correlation of client outcomes due to the attendance sequences, with the framework embedded in a hierarchical construction for modeling repeated measures data. One may use our approach to examine whether there is heterogeneity in the relative effectiveness of group therapy modules by identifying clusters of clients whose outcome trajectories vary across modules. Our framework is flexible enough to retain application-specific modeling choices. For the BRIGHT study, this includes specifying a proper conditionally autoregressive (CAR) base distribution for the non-parametric prior on module random

effects, which accounts for the open enrollment-induced client overlap in attendance of modules that are offered at adjacent time points (Paddock et al. 2011). We demonstrate that the DDP model may be re-cast for estimation as a DP under our multiple membership linkage of clients to treatment in a similar fashion as for the analysis of variance (ANOVA) DDP (De Iorio et al. 2004).

In §2, we introduce an additive model that employs client and MM random effects for BRIGHT study modules that was examined for open-enrollment group therapy data by Paddock & Savitsky (2012), and then build upon that work by introducing a multivariate generalization to allow for time-varying MM random effects. We present the DDP model in §3 to generalize the additive MM model to jointly model dependence owing to repeated measures within clients and group therapy module participation. Brief mention is made of our computational approach and software solution for conducting posterior simulations under the multiple membership models in §4, followed by an exploration of the properties of the models on simulated data in §5. Our motivating application focuses on the assessment of a group CBT intervention deployed in an open-enrollment study design for the treatment of depressive symptoms among clients in residential substance abuse treatment in §6. We conclude with discussion and conclusions in §7.

2 Multiple Membership Additive Semi-Parametric Models

This section introduces model constructions that include module random effects, which are mapped to each client according to the modules attended by that client using multiple membership modeling. These models permit inference about the rela-

tive effectiveness of the CBT intervention while accounting for differences in module effects as well as the dependence induced among clients based on overlaps in the sequences of modules attended. A separate client random effects term captures the within-client dependence among repeated measures.

2.1 Model Construction and Definitions

We first begin with the model of Paddock & Savitsky (2012) for modeling longitudinal post-treatment outcomes and allowing outcomes for clients who attend the same therapy group to be correlated:

$$y_{ij} = \mu + \mathbf{d}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{b}_i + \mathbf{x}_i^{'}\boldsymbol{\gamma} + \epsilon_{ij}, \tag{1}$$

where y_{ij} is the BDI-II depressive symptom score for client i ($i = 1, \ldots, n$) at repeated measurement event $j = (1, \ldots, o_i)$. The global intercept is represented by μ. $\mathbf{d}_{ij}$ are the fixed effects predictors and their associated effects are $\boldsymbol{\beta}$. We parameterize $\mathbf{d}_{ij} = \left(T_i, t_{ij}, t_{ij}^2, T_i t_{ij}, T_i t_{ij}^2\right)$ for the BRIGHT study, where T_i specifies an indicator for the treatment arm assigned to client i ($T_i = 1$ for clients receiving cognitive behavioral therapy (CBT), $T_i = 0$ for those receiving the "usual care" (UC)), and t_{ij} denotes the continuously-valued time at which y_{ij} was observed. The components of $\mathbf{d}_{ij}$ are chosen to estimate the effects on depressive symptom scores of CBT assignment, time, and the interaction of CBT assignment and time; a quadratic specification was chosen based on previous data analysis (Paddock & Savitsky 2012). The random effects predictor, $\mathbf{z}_{ij}$, is a $q \times 1$ vector associated with the q random effects for client i, $\{\mathbf{b}_i\}$. We set $\mathbf{z}_{ij} = (1, t_{ij}, t_{ij}^2)$ for the BRIGHT study, so that the $(q = 3) \times 1$ vector of random effect parameters for client i, $\mathbf{b}_i$, capture client-specific variation in change in BDI-II scores over time. Our parameterization of fixed and client random effects employs global second order polynomial terms to enforce smoothness and prevent

over-fitting under a study design with a relatively small number of measurement waves per client, as is typical of behavioral intervention studies such as BRIGHT. The second-to-last term allows for multiple membership modeling since depressive symptom scores observed post-treatment, y_{ij}, are not linked to specific therapy group modules, but rather to all modules attended by client i. This term maps the y_{ij}'s to the vector of S module random effects, $\boldsymbol{\gamma}$, by multiplying $\boldsymbol{\gamma}$ by an $S \times 1$ weight vector, $\mathbf{x}_i$, that is normalized to sum to 1 (Hill & Goldstein 1998). In particular, S_i equals the number of modules attended by client i; $x_{is} = 1/S_i$ if client i attended module s and $x_{is} = 0$ otherwise. Let $N = \sum_i o_i$ denote the number of repeated measures observed for all clients. Observational error is indicated by $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \tau_\epsilon^{-1}\right)$. We produce within-sample fitted client growth trajectories in §5 and §6 with employment of $(\boldsymbol{\beta}, \{\mathbf{b}_i\}, \boldsymbol{\gamma})$.

2.2 Distribution of Client Random Effects

Though one may parametrically model the client random effects, $\{\mathbf{b}_i\}$, we model them non-parametrically using a Dirichlet process (DP) prior to motivate the subsequent DDP development and to exploit the DP's usefulness for flexibly modeling the distribution of the $\{\mathbf{b}_i\}$'s despite having no more than three repeated measures per client in the BRIGHT study (Paddock & Savitsky 2012). Specify,

$$\mathbf{b}_1, \ldots, \mathbf{b}_n \big| F \overset{\text{iid}}{\sim} F \tag{2}$$

$$F \big| \alpha, \boldsymbol{\Lambda} \sim \text{DP}\left(\alpha, F_0\right), \tag{3}$$

where we choose base distribution, $F_0 \equiv \mathcal{N}_q\left(\mathbf{0}, \boldsymbol{\Lambda}^{-1}\right)$, a convenient conjugate form that spans the support for $\mathbf{b}$ and simplifies posterior sampling while still allowing the data to estimate a general form for F. We further specify, $\alpha \sim \mathcal{G}a\left(a_1 = 1, b_1 = 1\right)$, to allow the data to estimate the DP concentration parameter, reflecting its importance

for determining the total number of client clusters formed. We may equivalently enumerate (2) as a discrete mixture (Sethuraman 1994),

$$F = \sum_{h=1}^{\infty} p_h \delta_{\mathbf{b}_h^*}, \tag{4}$$

of countably infinite weighted point masses, where "locations" $(\mathbf{b}_1^*, \ldots, \mathbf{b}_M^*)$ index the unique values for the $\{\mathbf{b}_i\}$. The discrete construction for F allows for ties among sampled values for $\{\mathbf{b}_i\}$, so that $M \leq n$ and index clusters (i.e., clients sharing locations, or having same values of $\mathbf{b}$) with $n \times 1$, $\mathbf{s}$ where $s_i = m$ implies $\mathbf{b}_i = \mathbf{b}_m^*$. Then the set, $(\mathbf{s}, \{\mathbf{b}_m^*\})$, provides an equivalent parameterization to $\{\mathbf{b}_i\}$, though the former provides better mixing under posterior sampling (Neal 2000).

2.3 Distribution of Module Random Effects

2.3.1 Univariate Module Effects

Owing to the overlap in client attendance of modules under open enrollment into group therapy, we specify a conditionally autoregressive (CAR) prior for module random effects to allow them to be correlated. The degree of correlation is determined by the closeness of the modules, which depends on how we define which modules are neighbors. We define modules offered at adjacent time points within the same open-enrollment group as neighbors given that clients tend to attend subsequent modules in the BRIGHT study's residential treatment setting (Paddock et al. 2011).

To implement this, we enumerate a two-part form for the covariance matrix (Besag et al. 1991). Firstly, define an $S \times S$ adjacency matrix, $\mathbf{\Omega}$, to encode dependence among neighboring modules where we set $\omega_{ss'} = 1$ if module s is a *neighbor* of or "communicates" with module s' (denoted with "$\sim$" in $s \sim s'$), and 0 otherwise. Construct $\mathbf{D} = \text{Diag}(\omega_{s+})$, where $\omega_{s+} = \sum_j \omega_{sj}$ equals the number of neighbors of module s. Then compose the covariance matrix, $\mathbf{Q}^- = (\mathbf{D} - \mathbf{\Omega})^-$, the Moore-Penrose

pseudo-inverse, as Q is not of full rank, and specify the joint distribution of random module effects,

$$\boldsymbol{\gamma}|\ \tau_\gamma, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\tau_\gamma\left(\mathbf{D}-\boldsymbol{\Omega}\right)\right]^{-}\right), \tag{5}$$

where scalar precision parameter, τ_γ, controls the overall strength of variation. The rank of $(\mathbf{D}-\boldsymbol{\Omega})$ is $S-G$, where G represents the number of distinct open-enrollment therapy groups (Hodges et al. 2003). We use the following model short-hand label for simulated data and BRIGHT data analysis,

- **MMCAR:** Employ the additive model of Equation 1 under the joint prior construction of Equation 5 in the fashion of Paddock & Savitsky (2012).

Note that one could use a standard MM model for applications under which random effects may be assumed exchangeable.

2.3.2 Multivariate Module Effects

The univariate module effects may be replaced with a multivariate model specification that relaxes the assumption of constant module effects over time specified in Equation 5. Re-state Equation 1,

$$y_{ij} = \mu + \mathbf{d}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{b}_i + \left(\mathbf{x}_i^{'}\boldsymbol{\Gamma}\right)\mathbf{z}_{ij} + \epsilon_{ij}, \tag{6}$$

where $S \times q$, $\boldsymbol{\Gamma} = \left(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_S\right)^{'}$, for each of the multivariate $q \times 1$, $\boldsymbol{\gamma}_s$. We again assume a second order polynomial model, but this time for the module effects, where each module, s, is parameterized with a $(q=3)\times 1$ random effects vector back multiplied by $\mathbf{z}_{ij} = (1, t_{ij}, t_{ij}^2)$, which permits the effect of module s under the BRIGHT study to vary with time, t_{ij}. We may most easily make the extension of the CAR modeling of Besag et al. (1991) by stacking each of the q, $S \times 1$ columns from $\boldsymbol{\Gamma}$ into $qS \times 1$, $\mathfrak{G} = \left(\boldsymbol{\gamma}_{(1)}, \ldots, \boldsymbol{\gamma}_{(q)}\right)$ for the here $S \times 1$, $\boldsymbol{\gamma}_{(s)}$. Then compose the multivariate CAR prior,

$$\mathfrak{G}|\boldsymbol{\Lambda}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\left(\mathbf{D}-\boldsymbol{\Omega}\right)\otimes\boldsymbol{\Lambda}\right]^{-}\right), \tag{7}$$

for $qS \times qS$ precision matrix, $\mathbf{Q} = (\mathbf{D} - \mathbf{\Omega}) \otimes \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ describes the dependence among the q random effects per module and is specified to be identical to that used for base distribution associated with prior (2) imposed on $\{\mathbf{b}_i\}$. In summary, Equation 6 extends Equation 1 by permitting MM random (module) effects to vary over time. Assign the following label for our multivariate construction,

- **MM_MV:** Employ the additive model of Equation 6 under the joint prior construction of Equation 7.

2.4 Prior Distributions for Other Parameters

Scalar precision parameters, $(\tau_\epsilon, \tau_\gamma)$, are each specified with a $\mathcal{G}a(0.1, 0.1)$ prior with mean 1, while the $q \times q$ precision matrix, $\mathbf{\Lambda} \sim \mathcal{W}(q+1, \mathbb{I}_q)$, where the degrees of freedom are set to the minimum value to encourage updating by the data. Lastly, $(\mu, \boldsymbol{\beta})$, each receive non-informative priors. In instances where our priors specify fixed hyperparameters, we use values intended to be easily overwhelmed in the presence of data rather than eliciting them from our data.

3 Dependent Dirichlet Process for Multiple Membership Data

To allow for greater flexibility in modeling changes in module effects over time as well as the effects of modules on client depressive symptom trajectories, we now reformulate Equation 1 to explicitly index the client random effects by group therapy module identifiers, under which each client is assigned a $q \times (S+1)$ matrix of random effects. This contrasts with the previous specification of sets of $q \times 1$, $\{\mathbf{b}_i\}_{i=1,\ldots,n}$ client random effects and the $S \times 1$ module effects, $\boldsymbol{\gamma}$, given in Equation 1. The result-

ing client-by-module matrix parameterization arises from replacing a single random prior distribution for client effects with a collection of random prior distributions that are indexed by the unique module attendance sequences. First, we re-formulate Equation 1 in a more flexible composition,

$$
\begin{aligned}
y_{ij} &= \mu + \mathbf{d}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\boldsymbol{\Delta}_i\mathbf{x}_i + \epsilon_{ij} && (8)\\
\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_n \big| F &\overset{\text{iid}}{\sim} F && (9)\\
F \big| F_0 &\sim \text{DP}\left(\alpha, F_0\right), && (10)
\end{aligned}
$$

where we have replaced $q \times 1,\ \mathbf{b}_i$ and $S \times 1,\ \gamma$ with the $q \times (S+1),\ \boldsymbol{\Delta}_i$ for client i composed with,

$$
\boldsymbol{\Delta}_i = \left[\mathbf{b}_i, \mathbf{a}_{1,i}, \ldots, \mathbf{a}_{S,i}\right]. \tag{11}
$$

The first column of $\boldsymbol{\Delta}_i$ employs the analogous $\mathbf{b}_i$ client random effects from the additive models. The $\{\mathbf{a}_{s,i}\}_{s=1,\ldots,S}$ collect a set of $q \times 1$ module random effect vectors for client i. We note that every client receives an effect term, $\mathbf{a}_{s,i}$, for all of the S modules, even for modules they have not attended; such is even true for clients in the UC arm. By contrast, the additive model of Equation 1 is only defined at observed sequences of client module attendances, while this formulation is defined over a broader space of potential module attendance sequences across clients. We impose a DP prior on the set of client-by-module effects, $\boldsymbol{\Delta}_i$, in order that we may borrow strength and dimension reduce to discover clusters of clients expressing differential response sensitivities to treatment exposures. Employment of a continuous base distribution under the DP prior for the $\{\boldsymbol{\Delta}_i\}_{i=1,\ldots,n}$ allows the posterior inference on an arbitrary sequence of group therapy module linkages for each client. Effect values at unobserved modules are drawn from the non-degenerate continuous base distribution as updated by the *observed* module attendances. The module effect estimates for unobserved attendances for each client are set equal to the location values associated

with the cluster to which the client is assigned. The ability to develop a proper posterior distribution for arbitrary module attendance sequences is referred to by deIorio et al. (2004) as non-degeneracy.

Each of the $q \times 1$ columns of $\boldsymbol{\Delta}_i$ in Equation 8 is back multiplied by $\mathbf{x}_i$, which is the MM weight vector we earlier defined, but with a 1 prepended for a random intercept. More specifically, for $\mathbf{x}_i$ equal to some value $\mathbf{x}$, we construct the latter object as $\mathbf{x} \equiv (1, x_1, \ldots, x_S)$ for $x_s \in [0, 1]$ to encode the vector sequence for group therapy module attendance. Under our MM construction, the $(S+1) \times 1$, $\mathbf{x}$ is composed of values in $[0, 1]$ for $\sum_{s=1}^{S} x_s = 1$ for clients who attend at least one module, and $\sum_{s=1}^{S} x_s = 0$ for clients who do not.

We define the $q \times 1$ parameter vector, $\boldsymbol{\theta}_{\mathbf{x},i} \equiv \boldsymbol{\Delta}_i \mathbf{x}$, resulting from composition of the client-by-module random effects with the module attendance sequence. We write $\boldsymbol{\theta}_{\mathbf{x},i}$ and $\boldsymbol{\theta}_{\mathbf{x},i'}$ for clients i and i' that share the same attendance sequence, $\mathbf{x} \in \mathcal{X}$. Construct the subsequence, $\left(x_{s(1)}, \ldots, x_{s(K)}\right)$ for $K \leq S$ non-zero entries in $\mathbf{x}$ corresponding to modules attended for one or more clients with $\mathbf{x}_i = \mathbf{x}$. Then we may provide the more granular construction, $\boldsymbol{\theta}_{\mathbf{x},i} = \mathbf{b}_i + x_{s(1)}\mathbf{a}_{s(1),i} + \ldots + x_{s(K)}\mathbf{a}_{s(K),i}$, for client i where we note that only those modules attended by client i contribute to the likelihood. The multiplication of each $\mathbf{a}_{s(k)}$ by $x_{s(k)}$ reflects the MM design with $x_{s(k)} \in [0, 1]$.

Our formulation in Equation 8 may be re-expressed with the $q \times 1$ vector of client random effects, $\boldsymbol{\theta}_{\mathbf{x},i}$, in a similar fashion as the $q \times 1$ $\mathbf{b}_i$ in Equation 1, but here we index the client random effects by module attendance sequence, $\mathbf{x}$. The prior for $\boldsymbol{\theta}_{\mathbf{x},i}$ is specified under a collection of random distributions, $\{F_{\mathbf{x}}\}$, indexed by the unique attendance sequences, $\mathbf{x} \in \mathcal{X}$,

$$y_{ij} = \mu + \mathbf{d}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\boldsymbol{\theta}_{\mathbf{x},i} + \epsilon_{ij}, \tag{12}$$

with random effects vector, $\mathbf{z}_{ij}$, the same as composed in Equation 1. Specify the prior formulation for $\boldsymbol{\theta}_{\mathbf{x},i}$,

$$\boldsymbol{\theta}_{\mathbf{x},i}|F_{\mathbf{x}} \overset{\text{iid}}{\sim} F_{\mathbf{x}}. \tag{13}$$

We next enumerate a multiple membership dependent Dirichlet process (MM DDP) set of non-parametric distributions indexed by module attendance sequence, $\mathbf{x}$, in the stick-breaking construction (Sethuraman 1994),

$$F_{\mathbf{x}} = \sum_{h=1}^{\infty} p_h \delta_{\boldsymbol{\theta}^*_{\mathbf{x},h}}, \tag{14}$$

of weighted point mass locations where the weights are common for all values of $\mathbf{x} \in \mathcal{X}$, but the locations are indexed the unique attendance sequences (unlike for the simpler DP). We note that marginally, for each $\mathbf{x}$, the locations, $\boldsymbol{\theta}^*_{\mathbf{x},h}$ are exchangeable in h, such that $F_{\mathbf{x}}$ follows a Dirichlet process and we have established the propriety of the MM DDP. Denote the following short-hand notation for MM DDP construction,

$$\boldsymbol{\theta}_{\mathbf{x},i}|F_{\mathbf{x}} \overset{\text{iid}}{\sim} F_{\mathbf{x}} \tag{15}$$

$$\{F_{\mathbf{x}}, \mathbf{x} \in \mathcal{X}\} \sim \text{MM DDP}\left(\alpha, F_0\right), \tag{16}$$

where we have extended the ANOVA DDP prior of deIorio et al. (2004) to a multiple membership framework for the set of effect random distributions, $\{F_{\mathbf{x}}\}$.

We achieve Equation 8 from Equation 12 by extending a property of ANOVA DDP to the MM DDP that re-writes Equation 14 as a DP due to the finite indexing space of group therapy modules with,

$$F_{\mathbf{x}} = \sum_{h=1}^{\infty} p_h \delta_{\boldsymbol{\theta}^*_{\mathbf{x},h}} \tag{17}$$

$$= \sum_{h=1}^{\infty} p_h \delta_{\boldsymbol{\Delta}^*_h \mathbf{x}} = \delta_{\mathbf{x}} \sum_{h=1}^{\infty} p_h \delta_{\boldsymbol{\Delta}^*_h} \tag{18}$$

$$F_{\mathbf{x}} = \delta_{\mathbf{x}} F. \tag{19}$$

Then we may re-write our DDP model formulation of Equation 12 to the DP construction specified in Equation 8.

Though we use Equations 8 - 10 to estimate the MM DDP, the conceptual alternative in Equations 12-14 provides insight into the inferential properties of the MM DDP. The indexing of distributions, rather than just mean effects, by the module attendance sequences better spans the space of distributions generating the client random effects and allows the estimation of client module effects for modules not attended.

We also gain insight into the manner in which strength is borrowed over the set of module attendance sequences. The MM DDP formulation employs $\{F_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$ indexed by the set of unique module attendance sequences. Few clients, however, may be expected to exactly overlap or to share the same $\mathbf{x}$. Yet, clients will overlap for a portion of the module attendance sequences such that we have repeated observations for each module $s \in (1, \ldots, S)$ for estimation of the dependent $\{\mathbf{a}_{s,i'}\}_{i'}$ for all i' : $x_{s,i'} > 0$. The partial overlaps among the $\{\mathbf{x}\}_{\mathbf{x}\in\mathcal{X}}$ induce a dependence structure among the $\{F_{\mathbf{x}}\}$ based on the extent of overlaps.

3.1 Base Distribution

We structure the base distribution, F_0, for our $q\times(S+1)$ client-by-module parameters to leverage the adjacency dependence of the BRIGHT study modules. Compose F_0 for draws for the cluster locations, $\{\mathbf{\Delta}^*_m\}_{m=1,\ldots,M}$, as the product of multivariate Gaussian distributions for each of the $q\times 1$, $\mathbf{b}^*_m$ and the $q\times S$, $\mathbf{A}^*_m = [\mathbf{a}_{1,m}, \ldots, \mathbf{a}_{S,m}]$ that, together, comprise $\mathbf{\Delta}^*_m = [\mathbf{b}^*_m, \mathbf{A}^*_m]$ with,

$$\mathbf{b}^*_m \big| \mathbf{\Lambda} \overset{\text{iid}}{\sim} \mathcal{N}_q\left(\mathbf{0}, \mathbf{\Lambda}^{-1}\right) \tag{20}$$

$$\mathbf{A}^*_m \big| \mathbf{\Lambda}, \mathbf{\Omega}, \rho \overset{\text{iid}}{\sim} \mathbf{0} + \mathcal{N}_{q\times S}\left(\mathbf{\Lambda}^{-1}, \mathbf{Q}^{-1}\right), \tag{21}$$

where m indexes cluster location. The $\mathcal{N}_{q\times S}$ construction in Equation 21 employs a separable (parsimonious) covariance formulation for the distribution on the set of $q \times S$ matrix variate parameters, $\mathbf{A}_m^*$. We have employed the notation of Dawid (1981) under which the $q \times q$, $\mathbf{\Lambda}$, defines the precision matrix for the columns of $\{\mathbf{A}_m^*\}$ and the $S \times S$, $\mathbf{Q}$, for the rows. The covariance formulation is equivalent to $\mathrm{Cov}\left[\mathrm{vec}\left(\mathbf{A}_m^*\right)\right] = \mathbf{\Lambda}^{-1} \otimes \mathbf{Q}^{-1}$. (See Hoff (2011) for an intuitive discussion of separable covariance formulations.) Lastly, the preceding $\mathbf{0}$ presents the value of the $q \times S$ mean. Consistent with prior formulations under the additive models, the $q \times q$, $\mathbf{\Lambda} \sim \mathcal{W}(q+1, \mathbb{I}_q)$. We structure the $S \times S$ precision matrix, $\mathbf{Q}$, which models the module-induced adjacency dependence among the $q \times 1$ set, $\{\mathbf{a}_{m,s}^*\}_{s=1,\ldots,S}$, with a proper CAR formulation as enumerated in Jin et al. (2005), where $\mathbf{Q} = (\mathbf{D} - \rho\mathbf{\Omega})$, and $\rho \in (-1, 1)$ ensures $\mathbf{Q}$ is of full rank and may be viewed as a smoothing parameter that measures the strength of adjacency association. Matrices $(\mathbf{D}, \mathbf{\Omega})$ hold the same definitions as earlier specified in §2.3.

Proceeding with the notation of Dawid (1981), we pull together the components of the base distribution into,

$$F_0 = f\left(\mathbf{\Delta}_m^* \middle| \mathbf{\Lambda}, \mathbf{\Omega}, \rho\right) = \mathbf{0} + \mathcal{N}_{q\times(S+1)}\left(\mathbf{\Lambda}^{-1}, \mathbf{P}^{-1}\right), \tag{22}$$

where $\mathbf{P} = \mathrm{diag}\left(1, \mathbf{Q}\right)$. Let us prepare F_0 in the form we will use to conduct posterior simulations by stacking the q rows of $\mathbf{\Delta}_m^*$ (each an $(S+1) \times 1$ vector) to the $q(S+1) \times 1$, $\boldsymbol{\delta}_m^* = \left(\delta_{1,m}^{'}, \ldots, \delta_{q,m}^{'}\right)^{'}$ in,

$$F_0 = f\left(\boldsymbol{\delta}_m^* \middle| \mathbf{\Lambda}, \mathbf{\Omega}, \rho\right) = \mathcal{N}_{q(S+1)}\left(\mathbf{0}, \left[\mathbf{\Lambda} \otimes \mathbf{P}\right]^{-1}\right). \tag{23}$$

Vectorize $\mathbf{A}_m^*$ in a similar manner to obtain the $qS \times 1$, $\mathfrak{a}_m^* \middle| \mathbf{\Lambda}, \mathbf{\Omega}, \rho \overset{\text{iid}}{\sim} \mathcal{N}\left(\mathbf{0}, \left[(\mathbf{D} - \rho\mathbf{\Omega}) \otimes \mathbf{\Lambda}\right]^{-1}\right)$, which is similar to (7) but is full rank to permit efficient joint posterior sampling under high within-cluster dependence among the qS elements of $\mathfrak{a}_m^*$ under (9). Our MM

DDP formulation specifies the full set of S module effects for client i set equal to the location values, $\{\mathfrak{a}_m^*\}$, drawn from the CAR base distribution for cluster m that contains client i for some posterior sampling iteration.

Due to the BRIGHT study design, there were $G = 4$ open-enrollment therapy groups. Each group was composed of modules having at least partial overlap with another module with respect to the set of clients in attendance, and the sets of clients in the four groups were different. We thus add more flexibility in (23) by specializing the CAR prior in $\mathbf{P}$ to each open-enrollment therapy group with,

$$\mathbf{P} = \text{diag}\left(1, \mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_G\right), \tag{24}$$

where we have defined a set, $\{\mathbf{Q}_g\}_{g=1,\ldots,G}$, of CAR precision matrices composed as $S_g \times S_g$, $\mathbf{Q}_g = (\mathbf{D}_g - \rho_g \mathbf{\Omega}_g)$ and recover $\mathbf{D} = \text{diag}\left(\mathbf{D}_1, \ldots, \mathbf{D}_G\right)$ and $\mathbf{\Omega} = \text{diag}\left(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_G\right)$, reflecting the disjoint, non-communicating structure we seek to model. It is noted by Jin et al. (2005) that the parameterization of the global scalar smoothing parameter, ρ, may be overly restrictive, and they offer more heavily parameterized alternatives to permit the learning to adapt more locally. Our specification that offers the indexing of ρ_g by disjoint group allows smoothing across client-indexed module effects to be local to group. We may specify other continuous, multivariate distributions in place of the CAR for each group, including replacing the CAR covariance matrix construction with an anistropic Gaussian process (Savitsky & Vannucci 2010) or with an unspecified formulation under an inverse Wishart prior.

Assign the following label for the non-parametric construction,

- **DDP:** Equations 12-14 under the base distribution of Equation 23.

4 Computational Approach

Convergence of the sampler employed for simulation and the BRIGHT data analyses was assessed by employing a fixed width estimator with Monte Carlo standard errors (MCSE) computed using the consistent batch means (CBM) method (Jones et al. 2006). Computational software for the posterior distribution simulations is available in our package for the R statistical software (R Development Core Team 2011) package called `growcurves` (Savitsky & Paddock 2012). All of the methods, fit statistics and charts presented in this paper may be readily reproduced from `growcurves`. The parameters under DP priors are all sampled in a conjugate fashion by marginalizing over the random measure, F, to produce the Pólya urn scheme of Blackwell & MacQueen (1973), under which each cluster assignment indicator is sampled from a mixture of existing clusters and a new cluster. To the extent that a new cluster is selected, associated parameter locations are generated (and subsequently re-sampled) from the posterior of the base distribution under a single observation. (See Paddock & Savitsky (2012) for details).

We employ the cross-validatory, log pseudo marginal likelihood (LPML) leave-one-out fit statistic as described in Congdon (2005) under importance re-sampling of the posterior distributions over model parameters to estimate, $f\left(y_i|\mathbf{y}_{-i}, M_r\right)$, where M_r indexes our models where the leave-one-out property induces a penalty for model complexity and helps to assess the possibility for over-fitting. We also include the DIC_3 criterion of Celeux et al. (2006) that composes the marginal (predictive) density $\widehat{f\left(\mathbf{y}\right)}$ to estimate $f(\mathbf{y}|\theta)$ for composition of pD which is more appropriate for the (DP or DDP) mixture formulations that characterize all of our models. The non-penalized mean deviance, $\bar{\mathbf{D}}$, is also utilized.

5 Simulation Study

5.1 Data Generation

We generate data sets for simulation modeling from (12) by allocating the first 132 clients to the CBT and a remaining 168 to a non-group therapy usual care (UC) condition. We employ 24 modules for our simulation. Each CBT client attends 4 modules and each module on average holds 22 clients. The module attendance sequences, $\{\mathbf{x}_i\}$, used to select columns of the client-indexed matrix effects, $\{\mathbf{\Delta}_i\}$, are next generated in an open enrollment manner by randomly selecting the starting module for each CBT client in the block of 4 modules to which they are assigned. We set $\mathbf{x}_i = [\mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}]$ for all UC clients (who, by design, don't attend group therapy modules) as our hold-out or comparator module attendance sequence for identification. Such a design instantiates partial overlaps among the module attendance sequences for clients. The minimum and maximum numbers of clients linked to modules were restricted to 11 and 26, respectively, to conform to practical limitations on the underlying structure for group therapy modules. We simulate up to three repeated measures per client.

We simulate 4 clusters of clients, where each cluster generates a $(q = 3) \times ((S = 24) + 1)$ set of effect locations, $\mathbf{\Delta}^*_m$, shared by all clients assigned to them. The $q = 3$ rows of $\mathbf{\Delta}^*_m$ capture up to second order (intercept, linear and quadratic) polynomial effects for each module. The effects are generated in a vectorized fashion from a multivariate Gaussian with the covariance formulation as outlined for the DDP base distribution enumerated in §3.1. The module effects are generated from a multivariate proper CAR prior under the assumption of adjacency for successive modules with smoothing parameter, $\rho = 0.7$. A covariance matrix allowing for $q = 3$ polynomial

orders of module random effects is defined with,

$$\mathbf{\Lambda}^{-1} = \begin{bmatrix} 50 & -12 & 0.5 \\ -12 & 16 & -1.2 \\ 0.5 & -1.2 & 0.12 \end{bmatrix},$$

where the diagonals encode the variance of the first through third polynomial orders, respectively, for each of $(S = 24) \times 1$ multivariate cluster effect locations. We formulate $\mathbf{\Lambda}^{-1}$ such that the first and second orders and the second and third orders express negative correlations; for example, if the slope for the effect trajectory of a given module expresses a negative trajectory, then the quadratic term is positive and will tend to decelerate or bend the curve back up. Once the effects are generated, clients are randomly assigned to one of the 4 clusters with equal probability. Each cluster will hold both UC and CBT clients, though the module attendance sequence for the UC clients is set to 0's such that their assigned module effects do not contribute to the generation of the response values. The model intercept, μ, is set to 35 and fixed effect coefficients are set to $\beta = (-3, 0.25, 0, -2.5, 0.25)$ for $d_{ij} = (t_{ij}, t_{ij}^2, T_i, T_i t_{ij}, T_i t_{ij}^2)$, respectively, for each client, i, where T_i an indicator for the treatment arm assigned to client i ($T_i = 1$ for CBT, $T_i = 0$ for UC), and t_{ij} denotes the $j = 1, \ldots, 3$ continuously-valued time at which y_{ij} was observed, taking on value $0, 3$, or 6 months. The $q \times (S + 1)$ resultant set of random effects for client i, $\mathbf{\Delta}_i$, are multiplied with the $(S + 1) \times 1$ MM link vector, $\mathbf{x}_i$, to produce $q \times 1$, $\boldsymbol{\theta}_{\mathbf{x},i}$ matched to $\mathbf{z}_{ij} = (1, t_{ij}, t_{ij}^2)$ for client-specific polynomial variation from the mean time trend (which is captured in β). The model noise precision is set to $\tau_\epsilon = 0.1$.

5.2 Data Modeling

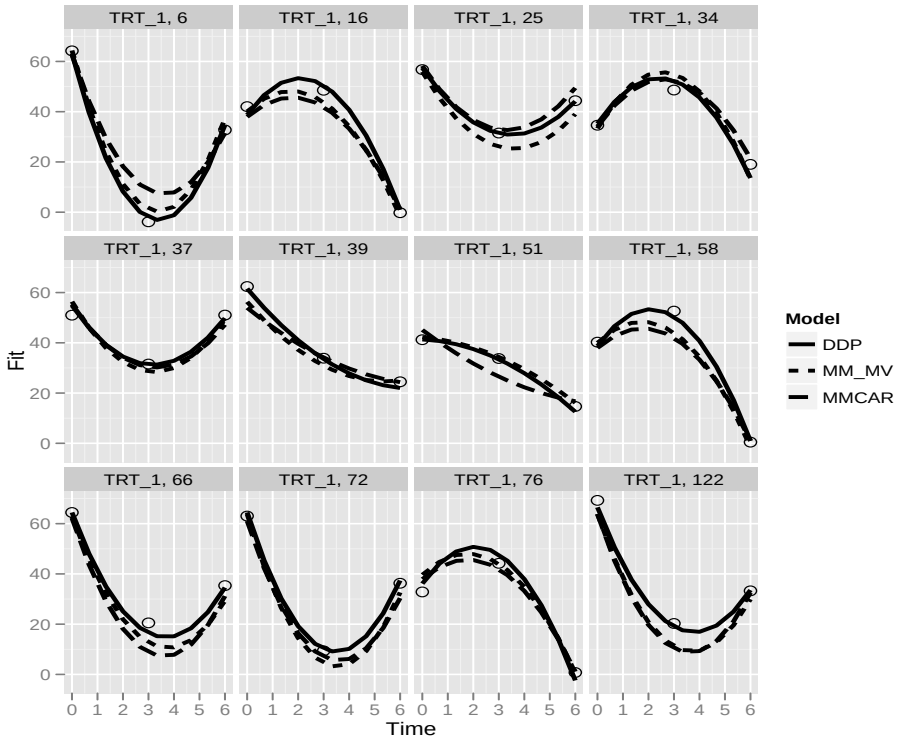

Figure 1: Posterior mean client growth curves under semi-parametric (MMCAR, MM_MV) and non-parametric (DDP) MM models under simulated data. Simulated data shown by circles.

Figure 1 presents in-sample predicted growth curves for randomly selected clients within each treatment arm along with actual client data values. Client growth curves under the DDP model express more adaptiveness to the data, both for U-shaped curves as expressed by client 6 and bell-shaped curves estimated for client 58.

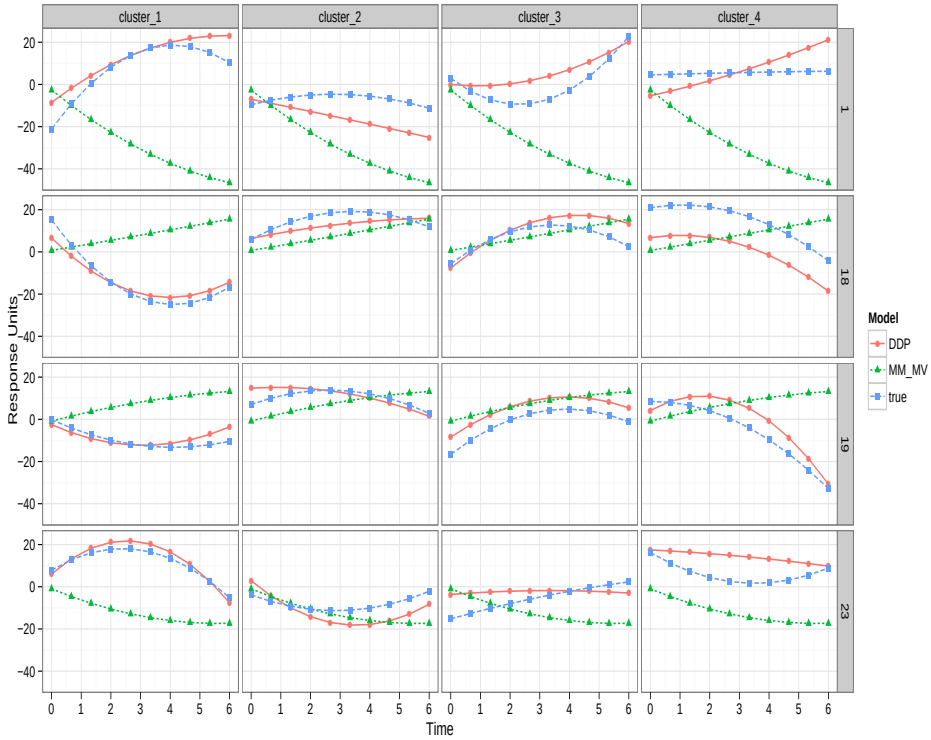

Figure 2: Posterior mean client module effect trajectories from simulated data for four clusters of clients, where the columns index clusters and the rows represent randomly selected modules. The curves are dimensioned in response units and represent the contribution of the modules to the response.

Posterior mean values for the 3 polynomial effect terms assigned to each module are composed into module effect trajectories through time in Figure 2 comparing MM_MV and DDP models for each of the 4 clusters (columns) and for 4 randomly-selected modules. The posterior mean module effect trajectories estimated under the DDP model track closer to the true trajectory shapes than do the non-client adaptive curves for MM_MV.

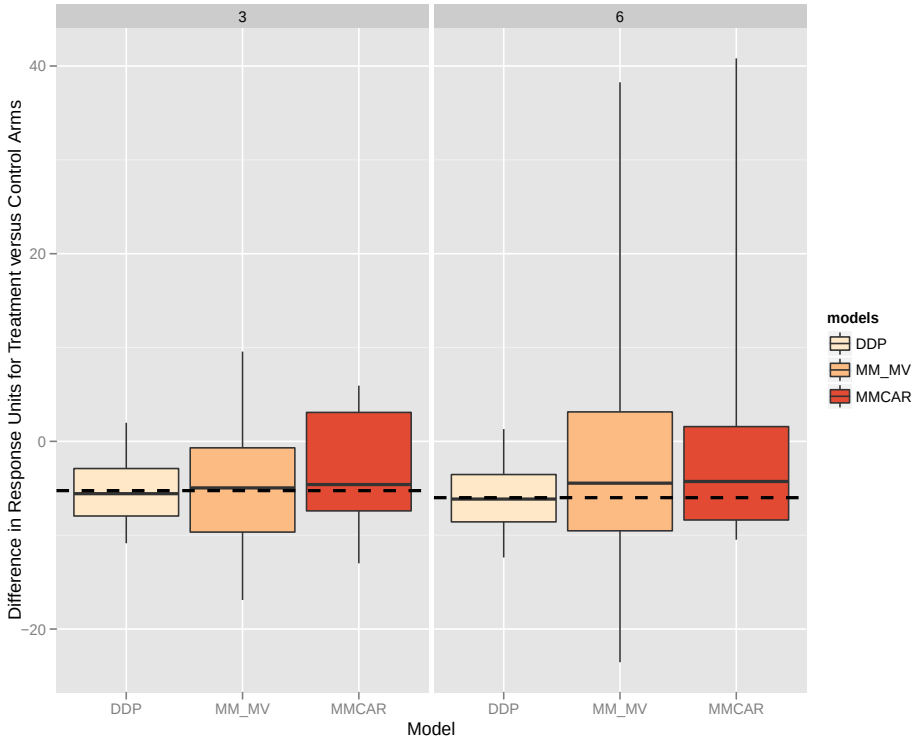

Figure 3: Predictive margins for the treatment effect of CBT versus usual care at 0 (left panel), 3 (middle panel) and 6 (right panel) months for MMCAR, MM_MV and DDP formulations under a Monte Carlo simulation from a data generating model where effects are indexed by cluster of clients and modules. Segments reflect the 95% credible intervals and boxes represent the interquartile range of the marginal posterior distribution. The dashed lines in each time period indicate the true treatment effects over the simulated datasets.

We compose a small Monte Carlo simulation with 10 iterations, where each generates a data set with the above-noted specifications. Estimation is performed under our models for each generated data set and the posterior draws for the fixed effects are concatenated across iterations to examine performance of the 3 comparator formulations under repeated sampling. Figure 3 reveals the posterior distribution over

the 95% credible intervals under each model estimated using the predictive margins technique; see Lane & Nelder (1982). We note that that the DDP formulation expresses the least uncertainty around the true values (represented by a dashed line at each of the 3 measurement months).

5.2.1 Model Fit Statistics

Model fit statistics, $\bar{D}$, $-LPML$, and DIC_3, are presented in Table 2. One observes lower (better) values across all 3 statistics for the DDP than the other two comparator models, while MM_MV, employing multivariate module effects, outperforms MMCAR parameterized with univariate module effects. In particular, the leave-one-out LPML statistic strongly prefers the DDP model. While the DDP is parameterized with client-by-module random effects, the effective parameterization is reduced under the clustering of clients. Nevertheless, the DDP would generally be expected to express a higher number of effective parameters than the two additive models, though the LPML performances do not indicate over-fitting. The polynomial construction for $\mathbf{z}_{ij}$ enforces smoothness in the estimated fit as demonstrated in the client growth curves from Figure 1, which also serves to mitigate the possibility for over-fitting. We performed additional simulations to explore scenarios 48 and 66 modules that, on average, have 11 and 8 clients per module, respectively, with the same number of clients. The relative model differences persist under $-LPML$. The $-LPML$ difference between DDP and MM_MV is 158 under $S = 24$ modules and 149 under 48 modules and 251 under 66 modules.

Table 1: Simulation Study Model Fit Comparisons: $\bar{\mathbf{D}}$, $-LPML$ and DIC_3 scores for model alternatives. Lower values imply better performance.

Model	$\bar{\mathbf{D}}$	$-LPML$	DIC_3
MMCAR	5073	2691	5208
MM_MV	4905	2592	5034
DDP	4607	2434	4715

6 Application to Group Therapy Data

We now return to the BRIGHT study for comparison of fit among our 3 model formulations. We further focus on inference under the MM DDP construction examine heterogeneity with respect to module type in BDI-II trajectories across disjoint clusters of clients. We recall our parameterization of fixed effects for the BRIGHT study data, $\mathbf{d}_{ij} = \left(T_i, t_{ij}, t_{ij}^2, T_i t_{ij}, T_i t_{ij}^2\right)$ where T_i an indicator for the treatment arm assigned to client i ($T_i = 1$ for CBT, $T_i = 0$ for UC), and t_{ij} denotes the continuously-valued time at which depressive symptom score, y_{ij}, was observed. As before, set $\mathbf{z}_{ij} = (1, t_{ij}, t_{ij}^2)$.

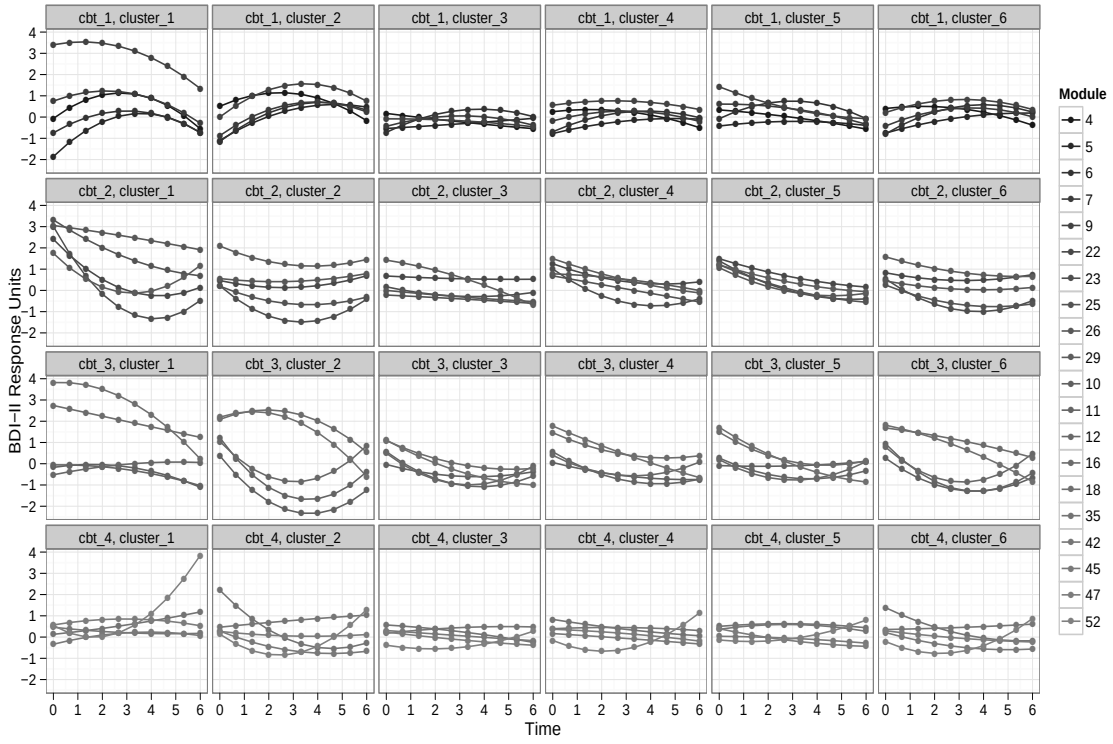

Figure 4: Posterior mean client module effect trajectories of BDI-II depressive symptom scores under the DDP model. Results are summarized by averaging trajectories into client clusters. Each plot cell is indexed by client cluster within each of 4 disjoint CBT groups, $(\text{cbt}_{-1}, \ldots, \text{cbt}_{-4})$. The rows of plot cells are indexed by CBT group and the columns, by cluster. The largest 6 clusters of clients are represented (in order of number of clients contained in each). Each plot contains module effect trajectories for randomly-selected modules within each of the four CBT groups.

We simplify and focus inference by composing posterior distributions for module effects up to clusters of clients. The client clustering is obtained from the among posterior samples of client partitions using the least squares algorithm of Dahl et al. (2008). The shapes, magnitudes and differences across the clusters express the range we see among clients so that we don't lose generality with a focus at the cluster,

rather than client, level. The most populated 6 clusters are employed and contain $(88, 51, 24, 23, 20, 19)$ clients, respectively, that together hold 225 out of 299 total BRIGHT study clients. Roughly half of the clients in the 6 clusters are UC clients who do not attend any group therapy modules. UC clients with mean client random effects, $\mathbf{b}_i$, similar to those of a subset of CBT clients are expected to co-cluster in posterior sampling such that the module effect values for all clients in the cluster are assigned the module effect location values for that cluster. This is an intuitive result where UC clients who express similar idiosyncratic characteristics to co-clustered CBT clients would be expected to similarly respond to CBT treatment were it offered to them.

Figure 4 renders module effect trajectories of the BDI-II depressive symptom scores for randomly selected modules. Results are summarized by averaging trajectories into client clusters, with the largest six clusters shown across the columns, denoted by $\text{cluster_1}, \ldots, \text{cluster_6}$. Each client cluster's trajectories are presented for each each of the 4 open-enrollment CBT therapy groups along the rows within clusters, which are denoted by $\text{cbt_1}, \ldots, \text{cbt_4}$ in the Figure. Large differences are observed in shape and magnitude among modules, particularly for client cluster 1, whose trajectories for each of the four open-enrollment groups are provided in the left-most column of plot cells of Figure 4. The range of the curves expresses clinically meaningful differences of $4 - 6$ (BDI-II) points (Furukawa 2010). Scanning the columns from left-to-right reveals a marked attenuation in cluster responsiveness to the CBT intervention. Member clients of clusters $4 - 6$ express much less depressive symptom sensitivity to participation in the modules and, therefore, one notes much less differentiation in effect values among the modules for these clusters.

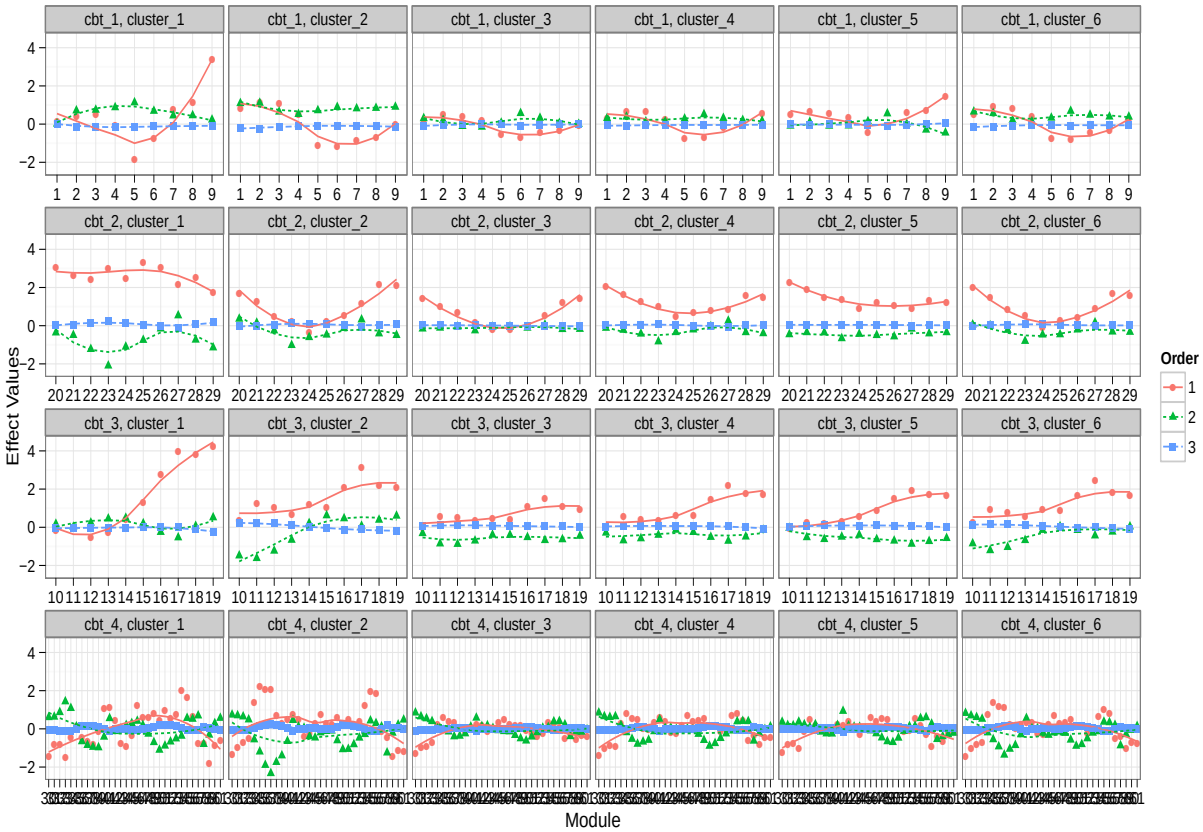

Figure 5: Posterior mean client module intercept (order 1), linear (order 2) and quadratic (order 3) effects for each module within the 4 disjoint CBT groups of modules averaged to cluster for BRIGHT case study under the DDP model. The rows of plot cells are indexed by CBT group and the columns, by cluster.

Figure 5 provides additional insight from the DDP model for examining the variation in module effects across clusters of clients and how those effects vary over time. The Figure shows module effect trajectories disaggregated into the $q = 3$ posterior mean polynomial effects from which they are each rendered across the 6 clusters of clients. The 3 polynomial effect values are presented for all modules, organized in the same cbt group-within-cluster format utilized in Figure 4. These polynomial parameters imply a module effect trajectory with the order 1 effect providing the intercept,

the order 2 effect the slope and order 3, a non-linear quadratic term. The resulting effects trajectory for a module would be U-shaped if the order 3 term is positive. As we noted in Figure 4, there is notable variation in the effect of modules on depressive symptoms across client clusters within each of the four CBT therapy groups as we scan from left-to-right, particularly for cbt groups $1-3$; for example, the first two clusters of each CBT therapy group, shown in the first two columns of Figure 5, show clinically meaningful variation in client outcomes.

Model fit statistics reveal an improved fit for the DDP in comparison to the other two models; however, unlike for the simulation results, the MMCAR produces a better fit for the BRIGHT data than does MM_MV. These results indicate the importance of differences across clients in responsiveness to modules. Within-sample predicted growth curves (not shown) demonstrate a similar improvement as observed in Figure 1 in shape and orientation adaptability for the DDP as compared to the other models to express better fit performance.

Table 2: BRIGHT Study Model Fit Comparisons: $\bar{D}$, $-LPML$ and DIC_3 scores for model alternatives. Lower values imply better performance.

Model	$\bar{\mathbf{D}}$	$-LPML$	DIC_3
MMCAR	5505	2980	5679
MM_MV	5547	2994	5716
DDP	5079	2929	5302

We explore sensitivity of the clustering of clients to our prior specification for the DP concentration parameter, α, employed in (10) for the MM DDP model by varying the shape and rate hyperparameters, (a_1, b_1), employed in the prior, $\alpha \sim \mathcal{G}a\,(a_1, b_1)$. We vary both hyperparameters in combinations within a range of $1-4$ for each,

producing a prior number of clusters from a minimum of 3 to maximum of 18. While our group therapy data application results show small differences in the posterior numbers of clusters formed, the allocation of clients to the most populous clusters is essentially unchanged as is our inference on client-module effects. Distributions for underlying parameters are also essentially unchanged.

7 Discussion

Our MM DDP approach extends the ANOVA DDP construction of De Iorio et al. (2004) to a multiple membership framework. The MM DDP provides wide support on the space of distributions indexed by the set of distinct multiple membership sequences through the borrowing of strength in overlaps among expressed sequences. The formulation allows one to examine whether element (e.g. module) effects vary across different client trajectories and vice-versa, allowing for one to learn about differing response sensitivities among clients to treatment elements, even for unobserved combinations of clients and treatment elements. We compose a model base distribution to retain straight-forward and efficient posterior sampling properties of the DP while allowing flexibility for Gaussian covariance specifications to parsimoniously parameterize dependence among module effects; in particular, we illustrate adjacency-based formulations for the covariance matrix of the Gaussian base measure in a fashion that renders flexibility while retaining conjugacy.

Other alternatives to our MM DDP may be considered, such as the hierarchical DP (HDP) (Teh et al. 2006) or the nested DP (NDP) (Rodríguez et al. 2008), which both target a grouped data structure with nested observations. These approaches, however, don't anticipate a multiple membership construction where sub-groups of clients share connections to the same modules as does the MM DDP, which indexes the

collection of random measures, $\{F_{\mathbf{x}}\}$, by multiple membership (attendance) sequence. While one may ignore the multiple membership composition and employ either of the HDP and NDP, they both perform posterior simulations in a nested, two-step, fashion (for a two-level hierarchical formulation), while we see how the MM DDP reduces to a DP that permits a simpler computational approach. Lastly, neither the HDP or NDP allow inference on unobserved module attendance sequences as does the MM DDP.

The usefulness of our approach may be limited for data with decreasing overlaps among the treatment element (e.g. client module attendance) sequences, $\{\mathbf{x}_i\}$, as this would restrict the ability for the data to borrow strength in the estimation of the collection of random distributions, $\{F_{\mathbf{x}}\}$. In one direction where clients perfectly overlap into disjoint groupings of client-modules for CBT studies, the MM DDP reduces to the ANOVA DDP. In the other direction, however, where clients express progressively less overlaps in modules attended, estimability may be compromised. In practice, resource limitations in the total number of modules offered for typical open-enrollment group therapy studies tend to produce a sufficient level of overlaps of clients on each module for estimation.

Software implementing the MM DDP is available for the R statistical software (R Development Core Team 2011) in a package called `growcurves` (Savitsky & Paddock 2012). All of the methods, fit statistics and charts presented in this paper may be reproduced from `growcurves`.

8 Acknowlegdements

This research was supported by the National Institute on Alcohol Abuse and Alcoholism grant to Susan Paddock (grant number R01AA019663). Data collection was

supported by National Institute on Alcohol Abuse and Alcoholism grant to Katherine Watkins (grant number R01AA014699).

References

Beck, A., Steer, R. & Brown, G. (1996), *Manual for the Beck Depression Inventory-II*, Psychological Corporation.

Besag, J., York, J. & Mollié, A. (1991), 'Bayesian image restoration, with two applications in spatial statistics (Disc: P21-59)', *Annals of the Institute of Statistical Mathematics* **43**, 1–20.

Blackwell, D. & MacQueen, J. B. (1973), 'Ferguson distributions via pólya urn schemes', *The Annals of Statistics* **1**, 353–355.

Carey, K. (2000), 'A multilevel modelling approach to analysis of patient costs under managed care', *Health Economics* **9**, 435–446.

Celeux, G., Forbes, F., Robert, C. P. & Titterington, D. M. (2006), 'Reply to comments on "Deviance information criteria for missing data models" (Pkg: P651-706)', *Bayesian Analysis* **1**(4), 701–706.

Congdon, P. (2005), *Bayesian models for categorical data*, Wiley series in probability and statistics, Wiley.

Crowe, T. & Grenyer, B. (2008), 'Is therapist alliance or whole group cohesion more influential in group psychotherapy outcomes?', *Clinical psychology and psychotherapy* **15**(4), 239–246.

Dahl, D. B., Day, R. & Tsai, J. W. (2008), 'Distance-based probability distribution on set partitions with applications to protein structure prediction', *Technical Report* .

Dawid, A. P. (1981), 'Some matrix-variate distribution theory: Notational considerations and a Bayesian application', *Biometrika* **68**, 265–274.

De Iorio, M., Müller, P., Rosner, G. L. & MacEachern, S. N. (2004), 'An ANOVA model for dependent random measures', *Journal of the American Statistical Association* **99**(465), 205–215.

deIorio, M., Müller, P., Rosner, G. L. & MacEachern, S. N. (2004), 'An anova model for dependent random measures', *Journal of American Statistical Association* **99**(2), 205–215.

Furukawa, T. (2010), 'Assessment of mood: Guides for clinicians', *Journal of Psychosomatic Reseach* **68**, 581–589.

Hill, P. W. & Goldstein, H. (1998), 'Multilevel modeling of educational data with cross-classification and missing identification for units', *Journal of Educational and Behavioral Statistics* **23**, 117–128.

Hodges, J. S., Carlin, B. P. & Fan, Q. (2003), 'On the precision of the conditionally autoregressive prior in spatial models', *Biometrics* **59**(2), 317–322.

Hoff, P. D. (2011), 'Separable covariance arrays via the Tucker product, with applications to multivariate relational data (Pkg: P179-208)', *Bayesian Analysis* **6**(2), 179–196.

Jin, X., Carlin, B. P. & Banerjee, S. (2005), 'Generalized hierarchical multivariate car models for areal data', *Biometrics* **61**(4), 950–961.

Jones, G. L., Haran, M., Caffo, B. S. & Neath, R. (2006), 'Fixed-width output analysis for Markov Chain Monte Carlo', *Journal of the American Statistical Association* **101**(476), 1537–1547.

Lane, P. W. & Nelder, J. A. (1982), 'Analysis of covariance and standardization as instances of prediction', *Biometrics* **38**, 613–621.

Langford, I. H., Leyland, A. H., Rasbash, J. & Goldstein, H. (1999), 'Multilevel modelling of the geographical distributions of diseases', *Journal of the Royal Statistical Society, Series C: Applied Statistics* **48**, 253–268.

Morgan-Lopez, A. & Fals-Stewart, W. (2006), 'Analytic complexities associated with group therapy in substance abuse treatment research: Problems, recommendations, and future directions', *Experimental and Clinical Psychopharmacology* **14**(2), 265–273.

Neal, R. M. (2000), 'Markov chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**(2), 249–265.

Paddock, S. M., Hunter, S. B., Watkins, K. E. & McCaffrey, D. F. (2011), 'Analysis of rolling group therapy data using conditionally autoregressive priors', *Annals of Applied Statistics* **5(2A)**, 605–627.

Paddock, S. M. & Savitsky, T. D. (2012), 'Bayesian hierarchical semiparametric modeling of longitudinal post-treatment outcomes from open-enrollment therapy groups', *To appear: Journal of the Royal Statistical Society, Series A: Statistics in Society* .

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

URL: *http://www.R-project.org/*

Rodríguez, A., Dunson, D. B. & Gelfand, A. E. (2008), 'The nested Dirichlet process', *Journal of the American Statistical Association* **103**(483), 1131–1154.

Ryum, T., Hagen, R., Nordahl, H., Vogel, P. & Stiles, T. (2009), 'Perceived group climate as a predictor of long-term outcome in a randomized controlled trial of cognitive-behavioural group therapy for patients with comorbid psychiatric disorders', *Behavioural and Cognitive Psychotherapy* **37**, 497–510.

Savitsky, T. D. & Vannucci, M. (2010), 'Spiked dirichlet process priors for generalized Gaussian process models', *Journal of Probability and Statistics* **2010**(Article ID 201489), 1 – 14.

Savitsky, T. & Paddock, S. (2012), *growcurves: Semiparametric Hierarchical Bayesian Modeling of Longitudinal Outcomes*. R package version 2.15.2.
URL: *http://CRAN.R-project.org/package=growcurves*

Sethuraman, J. (1994), 'A contructive definition of Dirichlet priors', *Stastica Sinica* **4**(2), 639–650.

Smokowski, P., Rose, S. & Bacallao, M. (2001), 'Damaging experiences in therapeutic groups: How vulnerable consumers become group casualties', *Small Group Research* **32**(2), 223–251.

Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. (2006), 'Hierarchical Dirichlet processes', *Journal of the American Statistical Association* **101**(476), 1566–1581.

Watkins, K. E., Hunter, S. B., Hepner, K. A., Paddock, S. M., de la Cruz, E., Zhou, A. J. & Gilmore, J. (2011), 'An effectiveness trial of group cognitive behavioral therapy for patients with persistent depressive symptoms in substance abuse treatment', *Archives of General Psychiatry* **68**(6), 1–8.