Published in final edited form as:

Stat Med. 2016 January 15; 35(1): 97–114. doi:10.1002/sim.6616.

Bayesian Restricted Spatial Regression for Examining Session Features and Patient Outcomes in Open-Enrollment Group Therapy Studies

Susan M. Paddock¹, Thomas J. Leininger^{1,2}, and Sarah B. Hunter¹

Susan M. Paddock: paddock@rand.org; Thomas J. Leininger: tj13@duke.edu; Sarah B. Hunter: shunter@rand.org

¹RAND Corporation

²Duke University

Abstract

Group-based interventions have been developed for treating patients across a range of health conditions. Enrollment into such groups often occurs on an open (or rolling) basis. Conditional autoregression modeling of random session effects has been proposed to account for the expected correlation in session effects associated with the overlap in patient participation session-to-session. However, when the analytic objective is to examine the relationship between a fixed-effect session feature and a patient outcome using conditional autoregression, confounding might arise if the fixed session feature of interest and the random session effects vary across sessions in similar ways, resulting in bias and inflated standard errors of a fixed-effect session feature of interest. Motivated by the goal of examining the relationships between outcomes and the session features of leader and session module theme, we applied restricted spatial regression to the analysis of patient outcomes collected from 132 participants in an open-enrollment group for treating depression among patients of a residential alcohol and other drug treatment program and adapted the approach to the multilevel data structure of open-enrollment group data. As compared to standard conditional autoregression, the restricted regression approach resulted in more precise estimates of regression coefficients of the module theme and leader predictor variables. The restricted regression approach provides an important analytic tool for group therapy researchers who are investigating the relationship between key components of open-enrollment group therapy interventions and patient outcomes.

Keywords

Conditional autoregression; group therapy; multilevel model; random effects; rolling admissions; spatial confounding

BACKGROUND

Group-based interventions have been developed to treat patients across a range of health conditions, including alcohol use disorders [1], depression [2], eating disorders [3], pain [4],

and cancer [5]. Enrollment into therapy groups often occurs on an open (or rolling) basis, with individual members entering and departing the group at different sessions. This is in contrast to closed-enrollment groups, for which membership in the group is constant session-to-session. Open-enrollment groups (OEGs) are employed in real-world health care delivery settings since they cost-effectively maximize the number of persons participating in therapy, shorten patient wait time for entrance into treatment, and help ensure a clinically necessary minimum number of members to maintain the group dynamic [1].

Understanding the effects of group therapy session-level features on participant outcomes is particularly important for OEGs, which typically undergo several important changes over time. The group leader is a session-level feature that might change over time and is of importance given the major role of group leaders in setting the group dynamic [6, 7]. Overt actions taken by the group leader might affect the group climate. These actions include addressing confidentiality or providing support to group members [8, 9], as well as more subtle characteristics such as perceived group leader intentions [10]. The material or content covered in OEG sessions is another session-level feature of interest. Open enrollment allows for the organization of material to be covered in a modular format of contiguously-offered blocks of sessions that cover thematically-connected material. The relationship between module themes offered at various sessions and patient outcomes is of interest to developers of behavioral therapy interventions who would like to improve the efficiency of these interventions by focusing on the most promising components [11, 12]. When entered as predictor variables into a regression model, these session-level features are regarded as fixed effects, whose values are assumed known and observed without error.

There has been increasing interest recently in how to properly analyze data arising from group-based interventions, driven by the concern that failing to account for the expected correlation of group participant outcomes could lead to under-estimation of the variance of model parameters, increasing the risk of Type I errors. In closed-group studies, modeling the correlation of group member outcomes is typically done by assuming random therapy group effects independently follow a common distribution [13, 14]. In contrast, for OEGs it is more appropriate to characterize the correlation among participant outcomes at the session level rather than the group level and to allow for random session effects to be correlated given the anticipated overlap in participant attendance session-to-session.

An analytic approach that achieves these objectives and improves model fit relative to ignoring session-to-session correlation in OEG studies is to model session random effects using conditional autoregression (CAR). The CAR approach is the first analytic approach applied to OEG data that is flexible enough to be used regardless of whether data were collected from participants as they attend sessions during the active treatment phase [15] or following treatment [16, 17]. The key insight motivating the use of CAR is the analogy between OEG sessions and geographic areas, with CAR modeling often applied to the latter to account for correlation in outcomes across geographic units [18, 19]. Similarly, the CAR approach allows one to exploit similarities, or borrow strength, across OEG sessions that are likely to have similar effects on outcomes owing to their closeness (or neighborliness). In the OEG context, spatial closeness of sessions could be defined based on their proximity to each other in time or the degree of overlap in participants shared between sessions [15].

When the analytic objective is to examine the relationship between a fixed session feature (e.g., group leader or module type) and a patient outcome using the CAR framework, confounding might arise if the fixed session feature of interest and the random session effects vary across sessions in similar ways. In such a case, the random session effects and the fixed session feature would compete to explain common variation across sessions. This could lead to biased regression coefficient estimates and unduly large variances for session features. Such spatial confounding has been noted in applications of CAR to geographically-referenced data [20]. Though we focus on CAR models given their use has been established for modeling OEG data, spatial confounding is not unique to CAR models and could arise when modeling data whenever there is residual spatial variation and a covariate with similar spatial variation [21]. Spatial confounding in OEG studies is potentially a concern, given that session features such as material covered in a session and group leader are often identical for sessions that are neighbors in some spatial sense – e.g., sessions offered consecutively or with overlapping patients.

In this paper, we examine the issue of confounding of fixed-effect session features of interest and session random effects in OEG studies. We illustrate the use of diagnostics for assessing the extent of spatial confounding [20] in the new application area of OEG studies. We present an application of the restricted spatial regression technique [22] to the analysis of patient outcomes collected from an OEG setting when the primary inferential focus is on the fixed-effect session features. We discuss how, in the absence of restricted spatial regression, the fixed and random session effects compete to explain common variation, which could distort estimates of the fixed-effect session feature regression coefficients and unduly inflate their posterior variances. We adapt the approach to accommodate the multilevel data structure of OEG data, and illustrate this approach with a motivating data set drawn from a trial of group cognitive behavioral therapy to treat depression among patients of a residential alcohol and other drug (AOD) treatment program.

MODELING APPROACH

Core multilevel model

Let m_i be the number of observations (sessions attended) for participant i , and $N = \sum m_i$ equal the number of observations across all participants $i=1, \dots, n$ and sessions $s=1, \dots, S$, where n and S are the number of participants and OEG sessions, respectively. The first stage of the multilevel model is:

$$Y = \mu * \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{X}_2 \mathbf{b} + \mathbf{U} \boldsymbol{\theta} + \boldsymbol{\varepsilon}. \quad (1)$$

In Equation (1), $\mathbf{Y}$ is an N -length vector for which each element, y_{is} , corresponds to an observation for participant i at session s . μ is the intercept term and $\mathbf{1}$ is an N -length vector of ones. $\mathbf{X}_1$ is a matrix of participant characteristics of dimension N -by- (k_1+1) , with each row corresponding to k_1 characteristics for participant i plus a term for t_{is} , the time (weeks), as of session s , since participant i entered the therapy group, and $\boldsymbol{\beta}$ is the (k_1+1) -length vector of regression coefficients corresponding to $\mathbf{X}_1$, with β_{k_1+1} the slope for time. The next term, $\mathbf{X}_2 \mathbf{b}$, is included to model variation in the trajectories of the measured outcome across participants by modeling participant-specific intercept and slope terms as random effects;

specifically, X_2 is an N -by- $2n$ design matrix, such that the row corresponding to an observation from participant i at session s has values $(1, t_{is})$ in columns $2i-1$ and $2i$, respectively, and zeros elsewhere, and the $2n$ -by-1 vector of participant-level random effects is $\mathbf{b} = (b_{01}, b_{11}, b_{02}, b_{12}, \dots, b_{0n}, b_{1n})$. $\mathbf{U}$ is an N -by- S matrix such that if row j corresponds to participant i attending session s , then $u_{js} = 1$ but equals 0 otherwise; $\boldsymbol{\theta}$ is a vector of length S with θ_s the effect of session s on y_{is} ; and $\boldsymbol{\varepsilon}$ is the observational error. The session-level stage of the multilevel model is:

$$\boldsymbol{\theta} = \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\gamma}, \quad (2)$$

where $\mathbf{W}$ is a matrix of dimension S -by- k_2 of the session features, $\boldsymbol{\alpha}$ is a vector of length k_2 representing the effects of $\mathbf{W}$ on $\boldsymbol{\theta}$; and $\boldsymbol{\gamma}$ is an S -by-1 the vector of random session effects, with γ_s the random effect for session s .

Conditional autoregression for session random effects

To account for the overlapping participant attendance session-to-session and the likely non-independence of random session effects, $\boldsymbol{\gamma}$ is modeled using conditional autoregression [15, 16]. Under CAR, $\boldsymbol{\gamma}$ has an improper multivariate normal distribution:

$$p(\boldsymbol{\gamma}|\tau) \propto \tau^{(S-G)/2} \exp\left(-\frac{\tau}{2} \boldsymbol{\gamma}' \mathbf{Q} \boldsymbol{\gamma}\right), \quad (3a)$$

centered at mean vector $\mathbf{0}$ and having precision matrix $\tau \mathbf{Q}$. $\mathbf{Q}$ is an S -by- S matrix with element q_{sj} reflecting the closeness of sessions s and j . $\mathbf{Q}$ is constructed as a function of a symmetric matrix, $\boldsymbol{\Omega}$, that defines which sessions are close (or ‘neighbors,’ in a spatial sense), with $\omega_{ss}=0$ by definition. Then, set $q_{sj} = -\omega_{sj}$ and $q_{ss} = \sum_j \omega_{sj} = \omega_{s+}$. When the scalar precision parameter, τ , is relatively large, the random effect for session s will be relatively similar to those for its neighboring sessions. G is the number of OEGs, which are sets of sessions that are disconnected in terms of participant attendance [23]. Even though (3a) is an improper prior distribution, the posterior distribution of $\boldsymbol{\gamma}$ will be proper if the prior distribution on τ is proper [19]. One might equivalently re-write (3a) as a series of proper prior conditional distributions:

$$\gamma_s | \boldsymbol{\gamma}_{-s} \sim N\left(\frac{\sum_{j \neq s} \omega_{sj} \gamma_j}{\omega_{s+}}, \frac{1}{\tau \omega_{s+}}\right). \quad (3b)$$

This formula explicitly details the smoothing, or borrowing of strength, that occurs in the session random effect estimation; the mean of the random effect for session s is a weighted average of the random session effects for session s ’s neighboring sessions, with the weights determined by the closeness of session s with its neighbors. The borrowing of strength is particularly useful here given the expected correlation of random effects for sessions in the same OEG and the typically relatively small number of participants per session in OEG studies. The conditional independence of each session’s random effect given its neighbors eases computational burden while still resulting in the desired unconditional dependence of session effects, given expected participant overlap session-to-session.

Spatial confounding diagnostics

Following Reich et al. [20] and Hodges and Reich [22], set $\mathbf{Q} = \mathbf{Z}\mathbf{D}\mathbf{Z}'$, the spectral decomposition of $\mathbf{Q}$, where $\mathbf{Z}$ is an orthogonal matrix of dimension S-by-S and $\mathbf{D}$ is also an S-by-S matrix that is diagonal with positive elements $d_{11} \dots d_{S-G, S-G} > 0$. The last G elements of $\text{diag}(\mathbf{D})$ are zero, which in the OEG context corresponds to the number of distinct OEGs reflected in the data. Equation (2) can be equivalently written as:

$$\boldsymbol{\theta} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{a}, \quad (4)$$

where $\mathbf{a} = \mathbf{Z}'\boldsymbol{\gamma}$ and is multivariately normally distributed with mean zero and precision matrix $\tau\mathbf{D}$:

$$\mathbf{a}|\tau \sim MVN(\mathbf{0}, \tau\mathbf{D}). \quad (5)$$

$\mathbf{Z}$ are canonical regressors of the random session effects, $\mathbf{a}$. The prior distribution for $\mathbf{a}$ is centered at $\mathbf{0}$. The matrix, $\mathbf{D}$, determines the degree to which estimation of the posterior mean of the $\mathbf{a}$'s will weight toward the prior mean of $\mathbf{0}$ given τ . The prior precision of a_s will be large if d_{ss} is large, resulting in a large prior weight on a_s 's prior mean of 0. In contrast, the prior precision of a_s will be small if d_{ss} is small, resulting in a low prior weight on a_s 's prior mean of 0. When d_{ss} is small, a_s acts more like a fixed effect than a random effect, which can be problematic when $\mathbf{W}$ and $\mathbf{Z}$ are highly correlated. In the extreme case that the prior precision of a_s is zero (i.e., $d_{ss}=0$), the s^{th} column of $\mathbf{Z}$, $\mathbf{Z}_s$, will act as a fixed effect for session s , leading in the usual way to collinearity with other fixed session effects in the model, $\mathbf{W}$. Thus, examining the correlation of each session feature, or column of $\mathbf{W}$, with each eigenvector, or column of $\mathbf{Z}$, is a useful diagnostic to detect spatial confounding.

Restricted random effects

The restricted regression technique [20, 22] could be used if there is a concern about correlation of $\mathbf{W}$ and $\mathbf{Z}$. The main idea is to restrict the random session effects to model variation in the space orthogonal to the fixed session effects, $\mathbf{W}$. This results in regression coefficients for $\mathbf{W}$ that are uncorrelated with the session random effects. To accomplish this, set:

$$\mathbf{P} = \mathbf{I}_S - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}',$$

where $\mathbf{P}$ is of dimension S-by-S and of rank $S-k_2$. Then re-write Equation (2) as:

$$\boldsymbol{\theta} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\gamma}, \quad (6)$$

where $\boldsymbol{\gamma}$ now models random variation in session effects in the space orthogonal to $\mathbf{W}$, and $\boldsymbol{\gamma}$ would be modeled using the CAR prior (Equation 3). Alternatively, since $\mathbf{P}$ is not of full rank, Equation (6) can be rewritten using a full-rank design matrix, $\mathbf{L}$, for the session random effects, where $\mathbf{L}$ is a matrix of dimension S-by- $(S-k_2)$ defined in Equation 11 of [22]:

$$\theta = W\alpha + L\gamma^{**}, \quad (7)$$

where $\gamma^{**} = L'\gamma$ is of length $(S-k_2)$. Under this model specification, γ^{**} can be modeled as a multivariate normal with mean vector 0_{S-k_2} and with precision matrix $\tau L'QL$.

MOTIVATING APPLICATION

GCBT study background

The motivating application is a community-based effectiveness trial of a group cognitive behavioral therapy (GCBT) intervention for treating residential AOD treatment patients having depressive symptoms [24]. The GCBT intervention tested in this study consisted of 16 two-hour sessions delivered twice weekly over the span of eight weeks, with sessions divided into four-session modules, each of which is focused on one of four themes: Thoughts, Activities, People and Substance Abuse. Participants entered the therapy group at the beginning of each of the four modules (i.e., every two weeks). Figure 1 illustrates how the 16 sessions of the GCBT are divided into modules of four sessions each and focus on distinct themes. Participants were able to enter the group at the first session of any of the modules. The group leader manuals that describe the four modules are available elsewhere [25].

The GCBT study employed a quasi-experimental design in which cohorts of participants at each of four study sites received either residential treatment as usual or residential treatment plus GCBT provided by trained AOD treatment counselors. Participants were assigned to receive either GCBT or treatment as usual according to whether GCBT was offered at their study sites at the time of their entry into residential AOD treatment. Participants were first screened for depression symptoms by residential staff using the Patient Health Questionnaire (PHQ-8) [26] 14 days after entering treatment. Fifty-nine percent of the participants screened at two weeks scored five or greater on the PHQ-8, corresponding to at least mild depression symptoms. The research team conducted a second screening to further determine eligibility. Inclusion criteria at the second stage were Beck Depression Inventory-II (BDI-II) scores greater than 17, indicative of moderate to severe depressive symptoms [27, 28]; the ability to speak and understand English; and receiving residential AOD treatment. Participants who screened positive for a self-reported bipolar disorder, schizophrenia, or cognitive impairment, or who were on federal probation or parole were excluded from the study.

Overall, 299 participants enrolled into the study, with 159 assigned to treatment as usual and 140 to GCBT, 132 of whom attended at least one GCBT session. For the purposes of this paper, we examined data from the 132 individuals who attended the GCBT sessions. Two hundred forty-five GCBT sessions were delivered over the course of the study. This included 14 cycles of the 16-session sequence (224 sessions), 20 additional sessions covering two module themes to accommodate those who joined the therapy group late for three particular 16-week sequences, and one additional session following a long holiday weekend owing to poor attendance of the regularly scheduled session. These 245 sessions were divided into four OEGs having distinct participants. Though participants were able to

initiate treatment at the first session of any module, they sometimes did not complete each module or all four modules as intended: 73% of participants attended at least half of the sessions (eight sessions), 45% attended 13 or more sessions, and 63% attended at least one session per module.

Participant-level data

Participant characteristics data on age, gender, and race/ethnicity were collected at screening. The PHQ-9 [29] was completed by GCBT intervention participants at the beginning of every other session, starting with the first session of each module, as well as at the last scheduled session for each participant. The PHQ-9 is a nine-item self-report measure that assesses the nine depression symptoms from the DSM-IV depression criteria. For each symptom, respondents are asked whether in the past two weeks he or she experienced that symptoms ‘not at all’ (corresponding to an item score of 0) to ‘nearly every day’ (corresponding to an item score of 3). The range of possible PHQ-9 scores is 0 through 27, with scores of 1–4 indicating minimal depression, 5–9 indicating mild depression, 10–14 indicating moderate depression, 15–19 indicating moderately severe depression, and 20–27 indicating severe depression. The PHQ-9 includes the same items as the PHQ-8 plus an item that asks about thoughts of self-harm or death [30]. Periodic completion of the PHQ-9 was part of GCBT to help group members increase their awareness of their mood. It also allowed group leaders to notice whether there have been changes in the severity of group members’ depression symptoms. Physical and mental health status using the Short Form-12 version 2 (SF-12v2) [31] and the Beck Depression Inventory (BDI-II) [27] were collected from participants at a baseline interview that occurred within two weeks of the first screening.

Session features

Previously reported analyses show GCBT leads to significantly greater reductions in post-treatment depressive symptoms than treatment as usual [2] and that depressive symptoms significantly decrease for GCBT participants during the active treatment phase [15]. However, some research questions remain about the association of GCBT features with depressive symptoms. One such question is whether module theme is associated with symptom scores [11, 12]. For example, some research suggests the behavioral activation component of CBT (e.g., the “Activities” module theme) is more effective than other components [32, 33].

Another important characteristic of group therapy that may vary session-to-session and potentially be related to patient outcomes is the group leader (i.e., the facilitator(s) of the group therapy session). All GCBT sessions offered in this study were facilitated by two co-leaders. A total of five different group leaders participated in the study. Figure 2 depicts which co-leaders delivered each session, with the vertical axis denoting group leaders 1 through 5 and the horizontal axis the 245 sessions; the four distinct OEGs are separated by the vertical lines near sessions 36, 76, and 116 in Figure 2. In addition to the major role group leaders have in setting the group dynamic, we are interested in whether there is variation among the group leaders in this particular study with respect to patient outcomes. Such an examination augments current strategies to examine measures of group leader fidelity to manualized interventions [34]. If there is variation in group leader effects in early

(e.g., Stage 1–2 [35]) trials under which group leaders are often carefully monitored, it might suggest the manualized intervention would not be tenable for more realistic settings or that further modifications are required. Understanding variation in group leader effects on outcomes in later-stage trials would provide insight into the natural variation possible in more naturalistic implementations of the treatment intervention.

ANALYSIS

To examine whether session-level features are associated with depressive symptoms, the PHQ-9 scores collected from GCBT participants during the active treatment phase are modeled using three approaches: (1) the CAR model of Equations 1–3; (2) the restricted CAR model of Equations 1–2 and 7; and (3) a longitudinal growth model (LGM), which is equal to Equation 1 with $\theta = W\alpha$. Under the LGM, the correlation of participant outcomes due to common OEG attendance is ignored. The LGM is included here given that some group therapy data analysts continue to ignore the correlation due to common participation by patients in group therapy settings [36] and also to provide a reference when examining the change in the posterior variances and posterior means of regression coefficients that results when applying CAR only (model CAR) versus CAR with restricted spatial regression (model RCAR). For CAR and RCAR, the distance matrix, Ω , is constructed by setting $\omega_{sj}=1$ if both $|s-j|=1$ and sessions s and j are in the same OEG and $\omega_{sj}=0$ otherwise.

In all three models, participant-level covariates described above are included in X_1 to control for any difference due to demographics [37] and physical and mental health. The matrix, W , contains $k_2=7$ columns that represent dummy variables representing module theme and leader; module theme is coded as a set of three dummy variables, and the ‘Thoughts’ module theme as the reference category, and group session leaders 1 through 4 are indicated by four dummy variables in W , with Leader 5 as the reference category. The random effects for participant i , $b_i=(b_{0i}, b_{1i})$, are assumed to follow a multivariate distribution with mean vector 0 and covariance matrix $diag(\sigma_0^2, \sigma_1^2)$. Patient-specific random intercept and slope terms accounted for correlated repeated measures within participant and allowed for heterogeneity in PHQ-9 score trajectories from the average trajectory. The observational error term, ϵ , is assumed to follow a normal distribution with mean vector 0 and covariance matrix $\sigma_\epsilon^2 I$.

The Bayesian approach requires us to specify prior distributions of the remaining model parameters. The intercept, μ , is given have a flat, improper prior to align with the CAR prior specification used in WinBUGS software (i.e., $p(\mu) \propto 1$). Our CAR model specification implies the intercept terms for the $G=4$ open-enrollment groups have a flat prior, though γ or γ^{**} could be reparameterized to make these intercept terms and their priors more explicit. Alternatively, Markov Chain Monte Carlo output could be post-processed to yield the implicit open-enrollment group fixed effects. β and α are given multivariate normal distributions with mean vectors 0 and covariance matrices $s_\beta^2 I$ and $s_\alpha^2 I$, respectively, with $s_\beta^2=s_\alpha^2=10$. The precision parameter corresponding to observational error is modeled as $\sigma_\epsilon^{-2} \sim \text{Gamma}(a_\epsilon, b_\epsilon)$, where the mean of X is a/b for the specification $X \sim \text{Gamma}(a, b)$, and with $a_\epsilon = b_\epsilon = 1$. In the absence of compelling rationale to guide the selection of priors

for β , α , and σ_{ϵ}^{-2} , we chose priors that are relatively vague, and we confirmed that analysis results are not sensitive to other hyperparameter choices. The prior on the precision parameter for the session random effects is modeled as $\tau \sim \text{Gamma}(a_{\gamma}, b_{\gamma})$, with $a_{\gamma} = 0.1$ and $b_{\gamma} = 0.2$. This specification implies an *a priori* precision of 1 for each session's effect in Equation 3b, since ω_{s+} in our analysis is typically equal to 2. This reflects an expectation that a session's random effect would deviate on average from that of its neighbors by 1 PHQ-9 point, but with the prior variable enough to allow for much larger or smaller deviations. This prior is appropriate given that a clinically meaningful reduction in PHQ-9 scores over a longer 6-month period is five points [38]. The priors placed on the precision terms for the participant-level random intercept and slope terms are $\sigma_0^{-2} \sim \text{Gamma}(a_0, b_0)$ and $\sigma_1^{-2} \sim \text{Gamma}(a_1, b_1)$, with $a_0 = a_1 = b_0 = b_1 = 0.1$. We conducted sensitivity analyses of these two prior choices by alternatively assuming *Uniform(0,20)* and *Uniform(0,5)* priors on the respective standard deviations terms, and confirming the results were insensitive to these choices. Specifically, we chose *Uniform(0,20)* for the former because we would not expect baseline PHQ-9 scores to have a larger standard deviation than 20, given the possible range of PHQ-9 scores, and we chose *Uniform(0,5)* for the latter since we would expect the random slope term to have a smaller standard deviation than the random intercept term. We further confirmed that conclusions drawn from the models were insensitive to the prior specification of (b_{0i}, b_{1i}) as uncorrelated.

Marginal posterior distributions of unknown parameters were obtained by iteratively sampling the conditional posterior distributions using Markov Chain Monte Carlo (MCMC) as implemented in WinBUGS Version 1.4.3 [39]. The number of MCMC iterations for each model was determined by using a relative fixed-width stopping rule [40]. This involved estimating the approximate sampling distribution for the Monte Carlo (MC) error of the posterior means and quantiles and then obtaining MCMC samples until the 95% confidence interval of the MC error for a target parameter was no larger than ζ times the posterior standard deviation of the target parameter, with the maximum ζ allowed to be 0.10. Since the study design provided for the collection of PHQ-9 scores only at every other session, PHQ-9 scores were assumed to be missing at random under the analysis model [41]. In order to track participant attendance session to session, these missing PHQ-9 scores were multiply imputed as draws from their posterior predictive distributions as implemented in WinBUGS.

RESULTS

Spatial confounding diagnostics results

Table 1 shows the range of correlations of each session dummy variable with the eigenvectors, $\mathbf{Z}$. The range of the correlations is substantial for each of the leader dummy variables, with correlations across all of the dummies ranging from -0.70 to 0.84 . This range indicates that several combinations of session module theme and random session effects might induce multicollinearity and variance inflation. The correlations between the session module theme dummy variables and $\mathbf{Z}$ are less strong, ranging from -0.39 to 0.41 . The magnitudes of the correlations suggest further investigation into potential confounding of fixed session features and random effects is warranted. Figure 3 shows the maximum correlation between the dummy variables for session fixed effects, $\mathbf{W}$, and the eigenvectors,

Z , on the x-axis, versus d_{ss} on the y-axis. The presence of a small d_{ss} when W and Z are highly correlated indicates there is a random effect that is functioning like a fixed effect in the model, resulting in confounding of the fixed session features and random session effects. There are several small values of d_{ss} when the maximum absolute value of the correlation between W and Z equals or exceeds 0.4, signaling a potential problem with confounding. Table 1 and Figure 3 together suggest that restricting the random effects to be in the space orthogonal to the session-level fixed effects is warranted.

Analysis of PHQ-9 scores

Table 2 shows the posterior means and the 95% posterior probability intervals for regression coefficients of individual-level covariates and session-level features as well as other contrasts between levels of the session module theme and leader terms resulting from the multivariable analyses of the PHQ-9 scores. The results in Table 2 are based on a large enough number of MCMC iterations to yield a maximum value of ζ equal to 0.10 [40]; however, 56% of the posterior means and quantiles shown in Table 2 are associated with a more stringent maximum value of $\zeta=0.02$ while an additional 37% with a maximum ζ of 0.05. Consistent with study hypotheses, PHQ-9 scores are decreasing in all models; the posterior mean coefficient of time is equal to -0.90 in the LGM and restricted CAR models and -0.91 in the CAR models, and in each case the upper end of the 95% posterior probability interval is less than 0. This result implies the expected decrease in PHQ-9 scores over the 8-week GCBT course is about 7 points. This exceeds a previously established clinically important level of change of 5 points, which was based on two standard errors of measurement [38]. None of the other predictors are associated with a 5-point difference in PHQ-9 scores. The rightmost two columns of Table 2 are the standardized differences between the posterior mean coefficients from the LGM versus CAR and LGM versus RCAR models, respectively. The standardized difference is calculated as the difference in the posterior means of the regression coefficient parameters for the two models divided by the posterior standard deviation (SD) of the same parameter from the LGM. As expected, the standardized differences of the RCAR versus LGM models tend to be smaller than those between the CAR and LGM. The difference between the two standardized difference columns are underlined when the columns differ by at least 0.50. The largest (underlined) standardized differences are observed for the leader coefficients. Inferences based on the regression coefficients about the associations of individual-level and session-level characteristics are the same across all three models for this particular analysis. However, the upper bound of the 95% posterior interval of the regression coefficient for the contrast between Leaders 2 and 5 shifts from -0.28 under LGM to -0.11 under the restricted CAR model, demonstrating the potential for different inferences to result under the models.

We computed the relative variance of a regression coefficient under CAR versus LGM as the posterior variance of the coefficient under CAR divided by the posterior variance under LGM. Table 3 shows the relative variance of the posterior distributions of regression coefficients and other contrast terms for the session module theme and leader. The first two columns compare the variance under the CAR model and restricted CAR model, respectively, to the LGM, which does not include session random effects. The variances across the models are essentially equal for the participant-level characteristics (e.g.,

demographics and SF-12), with the relative variances between 0.98 and 1.03, with ratios being less than 1 attributable to MCMC error. In contrast, the posterior variances under CAR for the leader regression coefficients are 1.57 to 1.82 times larger than those under LGM, and the relative posterior variances of the other leader contrast terms range from 1.32 to 2.05. The relative posterior variances of CAR versus LGM for the module theme coefficients and other module contrast terms are smaller, ranging from 1.22 to 1.36. The inflation in variance relative to LGM might be due to modeling the correlation of participant outcomes associated with common OEG attendance, to confounding of session fixed and random effects, or both. Examining the relative posterior variances of coefficients under restricted CAR versus LGM (middle column of each side of Table 3) provides insight into this question. In fact, the relative variances are lower for restricted CAR versus LGM. For the leader coefficients and other leader contrasts, the relative variances range from 1.06 to 1.15, and for the module theme 1.06 to 1.18, suggesting that most of the variance inflation in CAR versus LGM is due to confounding of session fixed and random effects. The third column of Table 3 shows the relative variance of CAR versus restricted CAR. The posterior variance of the time coefficient is 1.28 times higher under CAR versus LGM and is 1.16 times larger for restricted CAR versus LGM, with the posterior variance under CAR 1.10 times as large as that under restricted CAR. The posterior variance of the leader dummies and contrasts under the CAR model ranges from 1.33 to 1.97 times larger than under restricted CAR. This suggests that much of the variance inflation in the CAR model relative to the LGM model is attributable to the confounding of session fixed and random effects, the same confounding which is mitigated by the restricted regression model.

DISCUSSION

We present the use of restricted spatial regression to a new application area, the analysis of data from open-enrollment group therapy studies. We illustrate a case in which there is confounding between session fixed effects (i.e., group leader and therapy module type) and random session effects. The GCBT intervention exemplified here is characteristic of group therapy interventions delivered in real-life alcohol and other drug treatment settings, given the open (or semi-open) enrollment, modularized delivery, and changes in group leaders over time. The analytic concerns are also similar, such as accounting for the non-independence of participant outcomes due to the common attendance of sessions of the same therapy group by participants. As shown here, the goal of modeling the non-independence of participant outcomes using random session effects potentially conflicts with the goal of understanding the effect of fixed session features on participant outcomes. The restricted regression approach offers a way to partition the session-specific variance into distinct components in order to prioritize the investigation of specific session features of interest – in our case, of module theme and session leader - while still modeling the non-independence of participant outcomes with session random effects. Though we focus on the CAR model in this paper given its use has been established for modeling OEG data, spatial confounding is not unique to CAR models and could arise when modeling data whenever there is residual spatial variation and a covariate with similar spatial variation; confounding could similarly arise if residuals are temporally correlated and a covariate has similar temporal variation [21]. Thus, confounding of the type described here would need to be considered as a

possibility in future alternatives to the CAR model for OEG data that involve mixed effects modeling of session effects.

In terms of our substantive findings, we did not find meaningful variation in participant depression scores across different module themes or leaders. Variation in module theme effects might suggest that certain modules be given greater priority than others, while variation in leader effects would suggest that leaders could not consistently deliver the intervention in a real-life setting. Our findings about the lack of variation in group leader effects augment a previous finding that the GCBT group leaders delivered the sessions with high competence and adherence to the manual [34]. The GCBT study provided multi-day GCBT training and weekly supervision to all group leaders by licensed clinical psychologists to ensure consistent delivery; thus, a lack of significant group leader effects is not surprising. Our analysis provides an illustration of how one might use session-level predictors from an OEG study to examine such treatment implementation issues. The diagnostics we examined in this paper show the importance of assessing spatial confounding of fixed session features, such as module theme and leader, with session random effects, providing a new approach for researchers who are interested in examining session features in open-enrollment group therapy studies.

The restricted regression approach provides an important analytic tool for group therapy researchers who are investigating the relationship between key components of open-enrollment group therapy interventions and patient outcomes. Our analysis of data from an open-enrollment group therapy study illustrates that, in the absence of restricted spatial regression, the fixed and random session effects compete to explain common variation, which could distort estimates of the fixed-effect session feature regression coefficients and unduly inflate their posterior variances. Random session effects are primarily included in analysis models of participant data collected from open-enrollment groups so that the posterior variances of key model parameters reflect an adjusted (reduced) effective sample size given that participant outcomes are non-independent [22]. When the primary study hypotheses pertain to the fixed session effects as in the motivating example, it is appropriate to separate the common variation into that associated with the fixed versus random effects. The large relative variances of the posterior distribution of leader coefficients in particular highlight how essential the restricted regression approach was for these analyses. As our motivating example involved the analysis of data collected from participants during the active treatment phase, future analyses of post-treatment outcomes data would require the restricted regression approach to be embedded within a multiple membership-based CAR model data to account for the fact that each post-treatment datum is connected to not just one but to all sessions attended by a given participant [16, 42]. Future work could also involve using a new and computationally faster implementation of restricted regression that uses a reduced dimension basis for modeling the random effects in a way that disallows negative spatial correlations among random effects [43].

ACKNOWLEDGEMENTS

We thank the referees for helpful comments on the paper. We thank Kimberly Hepner and Suzanne Perry for their roles in data collection and Brett Ewing and Annie Zhou for data processing. The analysis was supported by Award Number R01AA019663 from the National Institute on Alcohol Abuse and Alcoholism. The content is solely the

responsibility of the authors and does not necessarily represent the official views of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) or the National Institutes of Health (NIH). Data collection was supported by NIAAA Grant R01AA014699.

APPENDIX: R and WinBUGS code

I. LGM model

```

model LGM
{
  ##### DATA INPUTS
  # y          : PHQ9 observation for case i (i.e, person[i] at time[i])
  # time       : a vector of observation times associated with y
  # person     : indexes the group therapy participants associated with y
  # Ncase      : number of PHQ-9 observations
  # Nperson    : number of group therapy participants in the data
  #
  ##### PARAMETERS
  # beta       : coefficients for X (indiv-level characteristics)
  # alpha      : coefficients for W (session-level characteristics)
  # prec       : observation-level precision (=1/sigma^2)
  # meanslope  : mean slope for PHQ-9
  # tau0inv    : precision of random growth intercept terms
  # tau1inv    : precision of random growth slope terms

  for(i in 1:Ncase)
  {
    y[i] ~ dnorm(mu[i], prec)
    mu[i] <- b0[person[i]] + (b1[person[i]]*time[i] +
      leader2[i]*alpha[1] + leader3[i]*alpha[2] +
      leader4[i]*alpha[3] + leader1[i]*alpha[4] +
      actions[i]*alpha[5] + people[i]*alpha[6] +
      substuse[i]*alpha[7])
  }

  for(i in 1:Nperson)
  {
    b0mean[i] <- alpha0 + female[i]*betapar[1] + age[i]*betapar[2] +
      race1[i]*betapar[3] + race2[i]*betapar[4] + race3[i]*betapar[5] +
      trans_ment_agg[i]*betapar[6] +
      trans_phys_agg[i]*betapar[7]
    # hierarchical centering of random person effects
    b1[i] ~ dnorm(meanslope, tau1inv)
    b0[i] ~ dnorm(b0mean[i], tau0inv)
  }
}

```

```

beta[1:7] ~ dmnorm(mu1.beta[], V1.beta[,]) # covariates X
alpha[1:7] ~ dmnorm(mu2.beta[], V2.beta[,]) # covariates W

# other leader contrasts
leader12 <- betapar[11] - betapar[8]
leader13 <- betapar[11] - betapar[9]
leader14 <- betapar[11] - betapar[10]
leader21 <- betapar[8]-betapar[11]
leader23 <- betapar[8]-betapar[9]
leader24 <- betapar[8]-betapar[10]
leader34 <- betapar[9]-betapar[10]

# other session module theme contrasts
actpeop <- betapar[12] - betapar[13]
actsubst <- betapar[12] - betapar[14]
peopsubst <- betapar[13] - betapar[14]

alpha0 ~ dflat()
meanslope ~ dnorm(0,0.1)
prec ~ dgamma(1,1)
tau0inv ~ dgamma(.1,.1)
taulinv ~ dgamma(.1,.1)
}

```

II. CAR model

```

model CAR
{
#### DATA INPUTS (IN ADDITION TO THOSE LISTED ABOVE IN LGM CODE)
# S: number of sessions
# num, adj, weights: inputs to car.normal function-see GeoBUGS manual
#       for full details. Briefly:
# num:    S-length vector with number of neighbors for each session.
#       num(i) = 1 if i is the first or last session on an open-
#       enrollment group and num(i) = 2 otherwise.
# adj: a vector indicating the sessions that are neighbors.
#       for example, adj = (2,1,3,2,4,...) indicates session 1 is
#       neighbors with 2, session 2 is neighbors with 1 & 3, session 3
#       is neighbors with sessions 2, and 4, etc.
# weights: a vector of 1's, or same length as adj

#### PARAMETERS (IN ADDITION TO THOSE LISTED ABOVE IN LGM CODE)
# gammavec: S-length random session effect vector for CAR model

```

```

# tauspatial: scalar precision term for session random effects

for(i in 1:Ncase)
{
  y[i] ~ dnorm(mu[i], prec)
  mu[i] <- b0[person[i]] + (b1[person[i]]*time[i] +
    leader2[i]*alpha[1] + leader3[i]*alpha[2] +
    leader4[i]*alpha[3] + leader1[i]*alpha[4] +
    actions[i]*alpha[5] + people[i]*alpha[6] +
    substuse[i]*alpha[7] + gammavec[session[i]])
}

for(i in 1:Nperson)
{
  b0mean[i] <- alpha0 + female[i]*betapar[1]+age[i]*betapar[2] +
    race1[i]*betapar[3]+race2[i]*betapar[4]+race3[i]*betapar[5] +
    trans_ment_agg[i]*betapar[6] + trans_phys_agg[i]*betapar[7]
  # hierarchical centering of random person effects
  b1[i] ~ dnorm(meanslope,taulinv)
  b0[i] ~ dnorm(b0mean[i], tau0inv)
}

gammavec[1:S] ~ car.normal(adj[], weights[], num[], tauspatial)

beta[1:7] ~ dmnorm(mu1.beta[], V1.beta[,]) # covariates X
alpha[1:7] ~ dmnorm(mu2.beta[], V2.beta[,]) # covariates W

# other leader contrasts
leader12 <- betapar[11] - betapar[8]
leader13 <- betapar[11] - betapar[9]
leader14 <- betapar[11] - betapar[10]
leader21 <- betapar[8]-betapar[11]
leader23 <- betapar[8]-betapar[9]
leader24 <- betapar[8]-betapar[10]
leader34 <- betapar[9]-betapar[10]

# other session module theme contrasts
actpeop <- betapar[12] - betapar[13]
actsubst <- betapar[12] - betapar[14]
peopsubst <- betapar[13] - betapar[14]

tauspatial ~ dgamma(precz1, precz2)

alpha0 ~ dflat()

```



```

meanslope ~ dnorm(0,0.1)
prec ~ dgamma(1,1)
tau0inv ~ dgamma(.1,.1)
taulinv ~ dgamma(.1,.1)
}

```

III. Restricted CAR model

The WinBUGS code requires input Qt22 and Lt, which are derived in R before calling WinBUGS.

R code to derive Qt22 and Lt

```

Wmtx = unique(cbind(newdat$session, newdat$leader1, newdat$leader2b,
newdat$leader3, newdat$leader4, newdat$actions, newdat$people,
newdat$substuse))
Wmtx = Wmtx[order(Wmtx[,1]),]
Wmtx = Wmtx[,-1]
Pcb = diag(245) - Wmtx%%solve(t(Wmtx)%*%Wmtx)%*%t(Wmtx)

# create adjacency matrix: sessions j and j+1 are adjacent if in the
same open-enrollment group

Q = matrix(0,245,245)
diag(Q) = numall # number of adjacent sessions
indic = cbind(1:244,2:245)
for(i in 1:35){
  Q[indic[i,1], indic[i,2]] = -1
  Q[indic[i,2], indic[i,1]] = -1
}

for(i in 37:75){
  Q[indic[i,1], indic[i,2]] = -1
  Q[indic[i,2], indic[i,1]] = -1
}

for(i in 77:115){
  Q[indic[i,1], indic[i,2]] = -1
  Q[indic[i,2], indic[i,1]] = -1
}

for(i in 117:244){
  Q[indic[i,1], indic[i,2]] = -1
  Q[indic[i,2], indic[i,1]] = -1
}

```

```
## Derive inputs Lt and Qt22 for Winbugs restricted CAR model
eq <- eigen(Q)
Z <- eq$vec
D <- eq$values
L <- t(eigen(Pcb)$vectors[,eigen(Pcb)$values>0.1e-12])
Qt22 <- L%*%Q%*%t(L)
Lt = t(L)
```

WinBUGS program

```
model RCAR
{
#### DATA INPUTS (IN ADDITION TO THOSE LISTED IN LGM AND CAR CODE)
# Nspat: equal to $S-K_2$, the length of gamma** from Equation 7
# Qt22: part of the precision matrix for gamma**, see above for derivation
# Lt: see R code above
#### PARAMETERS (SPECIFIC TO RCAR):
# Lgammaastast: S-length vector, rightmost term in Equation 7
# gammaastast: $S-K_2$-length vector, for full-rank model

  for(i in 1:Ncase)
  {
    y[i] ~ dnorm(mu[i], prec)
    mu[i] <- b0[person[i]] + (b1[person[i]]*time[i] +
      leader2[i]*alpha[1] + leader3[i]*alpha[2] +
      leader4[i]*alpha[3] + leader1[i]*alpha[4] +
      actions[i]*alpha[5] + people[i]*alpha[6] +
      substuse[i]*alpha[7] + Lgammaastast[session[i]]
  }

  for(i in 1:Nperson)
  {
    b0mean[i] <- alpha0 + female[i]*betapar[1] + age[i]*betapar[2] +
      race1[i]*betapar[3] + race2[i]*betapar[4] +
      race3[i]*betapar[5] + trans_ment_agg[i]*betapar[6] +
      trans_phys_agg[i]*betapar[7];
    # hierarchical centering of random person effects
    b1[i] ~ dnorm(meanslope,taulinv)
    b0[i] ~ dnorm(b0mean[i], tau0inv)
  }

  for(i in 1:S)
  {
```

```

    lgammaastast[i]<-inprod2(Lt[i,],gammaastast[])/sqrt(tauspatial)
    # divide by sqrt(tauspatial) to correct for not including
    # tauspatial in dmnorm call below to model gammaastast
  }

  beta[1:7] ~ dmnorm(mu1.beta[], V1.beta[],) # covariates X
  alpha[1:7] ~ dmnorm(mu2.beta[], V2.beta[],) # covariates W

  leader12 <- betapar[11] - betapar[8]
  leader13 <- betapar[11] - betapar[9]
  leader14 <- betapar[11] - betapar[10]
  leader21 <- betapar[8]-betapar[11]
  leader23 <- betapar[8]-betapar[9]
  leader24 <- betapar[8]-betapar[10]
  leader34 <- betapar[9]-betapar[10]

  actpeop <- betapar[12] - betapar[13]
  actsubst <- betapar[12] - betapar[14]
  peopsubst <- betapar[13] - betapar[14]

  tauspatial ~ dgamma(precz1, precz2)
  gammaastast[1:Nspat] ~ dmnorm(zero.vector[], Qt22[,])

  alpha0 ~ dflat()
  meanslope ~ dnorm(0,0.1)
  prec ~ dgamma(1,1)
  tau0inv ~ dgamma(.1,.1)
  tau1inv ~ dgamma(.1,.1)
}

```

REFERENCES

1. Monti, PM. Treating alcohol dependence : a coping skills training guide. 2nd edn.. New York: Guilford Press; 2002.
2. Watkins KE, Hunter SB, Hepner KA, Paddock SM, de la Cruz E, Zhou AJ, Gilmore J. An effectiveness trial of group cognitive behavioral therapy for patients with persistent depressive symptoms in substance abuse treatment. *Arch Gen Psychiatry*. 2011; 68:577–584. [PubMed: 21646576]
3. Tasca GA, Ramsay T, Corace K, Illing V, Bone M, Balfour L, Bissada H. Modelling longitudinal data from a rolling therapy group program with membership turnover: Does group culture affect individual Alliance? *Group Dynamics*. 2010; 14:151–162.
4. Mayer R. Answering the Call of Pain: Clinic Treats Both the Mental and Physical Manifestations of Pain. *Advance Healthcare Network for Physical Therapy and Rehab Medicine*. 2008; 19:10.
5. Rummans TA, Clark MM, Sloan JA, Frost MH, Bostwick JM, Atherton PJ, Johnson ME, Gamble G, Richardson J, Brown P, Martensen J, Miller J, Piderman K, Huschka M, Girardi J, Hanson J. Impacting Quality of Life for Patients With Advanced Cancer With a Structured Multidisciplinary Intervention: A Randomized Controlled Trial. *Journal of Clinical Oncology*. 2006; 24:635–642. [PubMed: 16446335]

6. Agazarian, YM.; Peters, R. The visible and invisible group. London: Karmac Books; 1981.
7. Yalom, ID. The theory and practice of group psychotherapy. Fourth edn. New York: Basic Books; 1995.
8. Smokowski PR, Rose S, Todar K, Reardon K. Postgroup-Casualty Status, Group Events, and Leader Behavior: An Early Look Into the Dynamics of Damaging Group Experiences. *Research on Social Work Practice*. 1999; 9:555–574.
9. Smokowski PR, Rose SD, Bacallao ML. Damaging experiences in therapeutic groups - How vulnerable consumers become group casualties. *Small Group Research*. 2001; 32:223–251.
10. Kivlighan DM, Tarrant JM. Does group climate mediate the group leadership-group member outcome relationship? A test of Yalom's hypotheses about leadership priorities. *Group Dynamics-Theory Research and Practice*. 2001; 5:220–234.
11. Drapkin ML, Tate SR, McQuaid JR, Brown SA. Does initial treatment focus influence outcomes for depressed substance abusers? *J Subst Abuse Treat*. 2008; 35:343–350. [PubMed: 18295436]
12. Paddock SM, Hunter SB. Does group cognitive behavioral therapy module type moderate depression symptom changes in substance abuse treatment clients? *Journal Of Substance Abuse Treatment*. 2014; 47:78–85. [PubMed: 24657006]
13. Lee KJ, Thompson SG. The use of random effects models to allow for clustering in individually randomized trials. *Clin Trials*. 2005; 2:163–173. [PubMed: 16279138]
14. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clin Trials*. 2005; 2:152–162. [PubMed: 16279137]
15. Paddock SM, Hunter SB, Watkins KE, McCaffrey DF. Analysis of Rolling Group Therapy Data Using Conditionally Autoregressive Priors. *Annals of Applied Statistics*. 2011; 5:605–627. [PubMed: 21857889]
16. Paddock SM, Savitsky TD. Bayesian Hierarchical Semiparametric Modelling of Longitudinal Post-treatment Outcomes from Open Enrolment Therapy Groups. *J R Stat Soc Ser A Stat Soc*. 2013; 176:795–808.
17. Savitsky TD, Paddock SM. Bayesian Semi- and Non-parametric Models for Longitudinal Data with Multiple Membership Effects in R. *Journal of Statistical Software*. 2014; 57:1–35. [PubMed: 25400517]
18. Besag J, York J, Mollie A. Bayesian image-restoration, with 2 applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991; 43:1–20.
19. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*. 2005; 14:35–59. [PubMed: 15690999]
20. Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*. 2006; 62:1197–1206. [PubMed: 17156295]
21. Paciorek CJ. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Stat Sci*. 2010; 25:107–125. [PubMed: 21528104]
22. Hodges JS, Reich BJ. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*. 2010; 64:325–334.
23. Lavine ML, Hodges JS. On rigorous specification of ICAR models. *The American Statistician*. 2012; 66:42–49.
24. Watkins KE, Hunter SB, Hepner KA, Paddock SM, de la Cruz E, Zhou AJ, Gilmore J. An effectiveness trial of group cognitive behavioral therapy for patients with persistent depressive symptoms in substance abuse treatment. *Arch Gen Psychiatry*. 2011; 68:577–584. [PubMed: 21646576]
25. Hepner KA, Miranda J, Woo S, Watkins KE, Lagomasino IT, Wiseman SH, Munoz R. Building Recovery by Improving Goals, Habits and Thoughts (BRIGHT): A Group Cognitive Behavioral Therapy for Depression in Clients with Co-Occurring Alcohol and Drug Use Problems -- Group Leader's Manual. TR-977/1-NIAAA, RAND Corporation. 2011
26. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA : the journal of the American Medical Association*. 1999; 282:1737–1744. [PubMed: 10568646]

27. Beck, AT.; Steer, RA.; Brown, GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
28. Buckley TC, Parker JD, Heggie J. A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *Journal Of Substance Abuse Treatment*. 2001; 20:197–204. [PubMed: 11516588]
29. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*. 2002; 32:509–515.
30. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A Systematic Review. *General hospital psychiatry*. 2010; 32:345–359. [PubMed: 20633738]
31. Ware, J.; Kosinski, M.; Turner-Bowler, D.; Gandek, B. How to score version 2 of the SF-12 health survey. Lincoln, RI: QualityMetric Incorporated; 2002.
32. Daughters SB, Magidson JF, Schuster RM, Safren SA. Act healthy: A combined cognitive-behavioral depression and medication adherence treatment for hiv-infected substance abusers. *Cogn Behav Pract*. 2010; 17:309–321. [PubMed: 21709737]
33. Ekers D, Richards D, Gilbody S. A meta-analysis of randomized trials of behavioural treatment of depression. *Psychology Medicine*. 2008; 38:611–623.
34. Hepner KA, Hunter SB, Paddock SM, Zhou AJ, Watkins KE. Training addiction counselors to implement CBT for depression. *Adm Policy Ment Health*. 2011; 38:313–323. [PubMed: 21626444]
35. Rounsaville BJ, Carroll KM, Onken LS. A Stage Model of Behavioral Therapies research: Getting started and moving on from stage I. *Clinical Psychology-Science and Practice*. 2001; 8:133–142.
36. Bauer DJ, Sterba SK, Hallfors DD. Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*. 2008; 43:210–236. [PubMed: 20396621]
37. Hunter SB, Paddock SM, Zhou A, Watkins KE, Hepner KA. Do client attributes moderate the effectiveness of a group cognitive behavioral therapy for depression in addiction treatment? *J Behav Health Serv Res*. 2013; 40:57–70. [PubMed: 22828976]
38. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring Depression Treatment Outcomes with the Patient Health Questionnaire-9. *Medical Care*. 2004; 42:1194–1201. [PubMed: 15550799]
39. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
40. Flegal JM, Gong L. Relative Fixed-Width Stopping Rules for Markov Chain Monte Carlo Simulations. *Statistica Sinica*. 2015 to appear. Also available on *arXiv preprint arXiv:1303.0238*.
41. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. [PubMed: 12090408]
42. Savitsky TD, Paddock SM. Bayesian Non-Parametric Hierarchical Modeling for Multiple Membership Data in Grouped Attendance Interventions. *Ann Appl Stat*. 2013; 7:1074–1094.
43. Hughes J, Haran M. Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2013; 75:139–159.

Figure 1.
Delivery of the 16-session course of GCBT by session theme (module)

Figure 2.
Co-leaders for each GCBT session

Figure 3.

The maximum absolute correlation between the dummy variables for session features, W , and the eigenvectors, Z (horizontal axis), versus d_{ss} (vertical axis) for each session s , $s=1, \dots, 245$.

Table 1
Range of correlations for each fixed session effect dummy variable with the eigenvectors, Z

	Group Leader Dummy Variables				Module Theme Dummy Variables		
	<i>Leader 1</i>	<i>Leader 2</i>	<i>Leader 3</i>	<i>Leader 4</i>	<i>Activities</i>	<i>People</i>	<i>Substance Abuse</i>
Minimum	-0.51	-0.70	-0.46	-0.52	-0.39	-0.37	-0.27
Mean	0.00	0.00	0.00	0.01	-0.01	0.00	0.00
Maximum	0.84	0.49	0.38	0.59	0.22	0.28	0.41

Table 2

Posterior means and 95% posterior probability intervals for regression coefficients and other contrast terms for session fixed effects in the model of PHQ-9 under the longitudinal growth model without session random effects (LGM), conditional autoregression (CAR), and restricted CAR (RCAR)

	LGM			CAR			RCAR			Standardized difference between means from LGM and * :
	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	CAR	RCAR
Regression Coefficients										
Time (weeks)	-0.90	(-1.02, -0.78)	-0.91	(-1.05, -0.78)	-0.91	(-1.03, -0.78)	-0.17	(-1.03, -0.78)	-0.17	-0.17
Female	-0.70	(-2.20, 0.81)	-0.91	(-2.43, 0.63)	-0.74	(-2.26, 0.78)	-0.27	(-2.26, 0.78)	-0.27	-0.05
Age	0.18	(-0.66, 1.01)	0.26	(-0.59, 1.11)	0.21	(-0.63, 1.06)	0.19	(-0.63, 1.06)	0.19	0.07
Race/ethnicity (vs. White)										
Black	-0.24	(-1.99, 1.51)	0.07	(-1.70, 1.85)	-0.15	(-1.91, 1.61)	0.35	(-1.91, 1.61)	0.35	0.10
Hispanic	-0.23	(-2.14, 1.68)	-0.19	(-2.11, 1.744)	-0.14	(-2.06, 1.78)	0.04	(-2.06, 1.78)	0.04	0.09
Other	-0.40	(-2.76, 1.96)	-0.43	(-2.78, 1.92)	-0.40	(-2.74, 1.94)	-0.02	(-2.74, 1.94)	-0.02	-0.00
SF12 Mental	-0.14	(-0.21, -0.07)	-0.15	(-0.22, -0.08)	-0.15	(-0.22, -0.08)	-0.28	(-0.22, -0.08)	-0.28	-0.28
SF12 Physical	-0.08	(-0.15, -0.02)	-0.09	(-0.15, -0.03)	-0.09	(-0.15, -0.02)	-0.31	(-0.15, -0.02)	-0.31	-0.31
Session leader (vs. Leader 5):										
Leader 1	-0.76	(-2.40, 0.89)	-0.74	(-2.82, 1.35)	-0.81	(-2.55, 0.95)	0.02	(-2.55, 0.95)	0.02	-0.06
Leader 2	-1.51	(-2.73, -0.28)	-1.79	(-3.32, -0.25)	-1.44	(-2.77, -0.11)	-0.45	(-2.77, -0.11)	-0.45	0.11
Leader 3	-0.26	(-1.29, 0.76)	-0.58	(-1.92, 0.76)	-0.26	(-1.36, 0.83)	-0.61	(-1.36, 0.83)	-0.61	0.00
Leader 4	-0.68	(-1.82, 0.46)	-0.81	(-2.36, 0.73)	-0.86	(-2.06, 0.35)	-0.22	(-2.06, 0.35)	-0.22	-0.31
Module theme (vs. Thoughts)										
Activities	-0.47	(-1.11, 0.16)	-0.42	(-1.12, 0.28)	-0.39	(-1.05, 0.27)	0.16	(-1.05, 0.27)	0.16	0.25
People	-0.39	(-1.07, 0.29)	-0.46	(-1.24, 0.32)	-0.36	(-1.07, 0.35)	-0.20	(-1.07, 0.35)	-0.20	0.09
Substance Abuse	-0.17	(-0.86, 0.52)	-0.25	(-1.02, 0.51)	-0.15	(-0.89, 0.59)	-0.20	(-0.89, 0.59)	-0.20	0.09
Other contrasts:										
Other leader contrasts:										
Leaders 1 vs. 2	0.75	(-0.99, 2.49)	1.05	(-1.41, 3.53)	0.64	(-1.15, 2.44)	0.34	(-1.15, 2.44)	0.34	-0.12
Leaders 1 vs. 3	-0.49	(-1.94, 0.95)	-0.16	(-2.20, 1.88)	-0.54	(-2.03, 0.96)	0.45	(-2.03, 0.96)	0.45	-0.07
Leaders 1 vs. 4	-0.08	(-1.73, 1.58)	0.07	(-2.31, 2.44)	0.06	(-1.65, 1.76)	0.18	(-1.65, 1.76)	0.18	0.17

	LGM			CAR			RCAR			Standardized difference between means from LGM and * :
	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)	CAR	RCAR
Leaders 2 vs. 3	-1.24	(-2.53, 0.05)	-1.21	(-2.88, 0.48)	-1.18	(-2.53, 0.17)	0.05	0.09		
Leaders 2 vs. 4	-0.82	(-2.04, 0.39)	-0.98	(-2.74, 0.84)	-0.58	(-1.86, 0.69)	<u>-0.26</u>	<u>0.39</u>		
Leaders 3 vs. 4	0.42	(-0.73, 1.57)	0.23	(-1.08, 1.55)	0.60	(-0.61, 1.80)	<u>-0.32</u>	<u>0.31</u>		
Other module theme contrasts:										
Activities vs. People	-0.08	(-0.76, 0.59)	0.04	(-0.71, 0.79)	-0.03	(-0.72, 0.66)	0.35	0.15		
Activities vs. Substance Abuse	-0.30	(-0.99, 0.38)	-0.17	(-0.97, 0.64)	-0.24	(-0.96, 0.48)	0.37	0.17		
People vs. Substance Abuse	-0.22	(-0.90, 0.46)	-0.21	(-0.97, 0.55)	-0.21	(-0.90, 0.49)	0.03	0.03		

* A standardized difference equals the difference in posterior means between CAR (or RCAR) versus LGM, divided by the posterior standard deviation under the LGM. When the standardized differences of CAR and RCAR (vs. LGM) differ by more than 0.5, these are underlined.

Table 3

Relative variance of the posterior distribution of regression coefficients and contrast terms for session fixed effects under CAR and restricted CAR

	Relative variance under:			Relative variance under:		
	CAR vs. LGM	restricted CAR vs. LGM	CAR vs. restricted CAR	CAR vs. LGM	restricted CAR vs. LGM	CAR vs. restricted CAR
Regression coefficients						
Time (weeks)	1.28	1.16	1.10			
Female	1.03	1.02	1.01			
Age	1.03	1.03	1.00			
Race/ethnicity						
(vs. White)						
Black	1.03	1.01	1.02			
Hispanic	1.01	1.00	0.99			
Other	0.99	0.98	1.01			
SF12 Mental	1.01	1.00	1.01			
SF12 Physical	0.99	0.99	1.00			
Session Leader						
(vs. Leader 5):						
Leader 1	1.60	1.12	1.42			
Leader 2	1.57	1.18	1.33			
Leader 3	1.71	1.14	1.49			
Leader 4	1.82	1.12	1.62			
Module theme (vs. Thoughts)						
Activities	1.23	1.09	1.13			
People	1.31	1.08	1.22			
Substance Abuse	1.22	1.15	1.06			
Other Contrasts:						
Other Leader contrasts:						
Leaders 1 vs. 2				2.02	1.06	1.89
Leaders 1 vs. 3				1.99	1.07	1.86
Leaders 1 vs. 4				2.05	1.06	1.94
Leaders 2 vs. 3				1.71	1.10	1.55
Leaders 2 vs. 4				2.16	1.10	1.97
Leaders 3 vs. 4				1.32	1.10	1.19
Other module theme contrasts:						
Activities vs. People				1.24	1.06	1.16
Activities vs. Substance				1.36	1.09	1.24
Abuse vs. People vs. Substance						
Substance Abuse				1.24	1.04	1.19