

Efficient Bayesian joint models for group randomized trials with multiple observation times and multiple outcomes

Xinyi Xu,^a Michael L. Pennell,^{b,*†} Bo Lu^b and David M. Murray^c

In this paper, we propose a Bayesian method for group randomized trials with multiple observation times and multiple outcomes of different types. We jointly model these outcomes using latent multivariate normal linear regression, which allows treatment effects to change with time and accounts for (i) intraclass correlation within groups; (ii) the correlation between different outcomes measured on the same subject; and (iii) the over-time correlation of each outcome. Moreover, we develop a set of innovative priors for the variance components, which yield direct inference on the correlations, avoid undesirable constraints, and allow utilization of information from previous studies. We illustrate through simulations that our model can improve estimation efficiency (lower posterior standard deviations) of intraclass correlations and treatment effects relative to single outcome models and models with diffuse priors on the variance components. We also demonstrate the methodology using body composition data collected in the Trial of Activity in Adolescent Girls. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: Bayesian methodology; group randomized trial (GRT); intraclass correlation (ICC); joint modeling; prior elicitation

1. Introduction

In many public health studies, the goal is to evaluate interventions intended to improve the health of communities through education and promotion of healthy behaviors. For example, some studies have examined interventions to increase physical activity levels in a school [1] or a workplace [2], and some have aimed to increase cancer screening rates in underserved populations [3]. The cluster or group randomized trial (GRT) is regarded as the gold standard for evaluating interventions delivered to communities [4, 5]. Groups, rather than individuals, are randomized to each treatment arm with randomization potentially stratified by factors believed to affect the outcome variable. In community-based research, GRTs are usually the only way to test an intervention without risk of bias owing to contamination, which occurs when important components of the intervention find their way into the control condition. GRTs are also common in primary care research in which the goal is to change physician behavior and patient management in hopes of improving patient outcomes, such as secondary prevention of coronary heart disease [6].

The statistical implications of GRTs were first brought to light by Cornfield [7]. He identified two important penalties of this experimental design. First, members within the same group often have physical, social, or geographical connections with each other, so their responses are likely correlated. This correlation is called the intraclass correlation (ICC), and it is positive in most situations. Failure to account for the ICC in analysis can result in inflated type-I error [8–10], and failure to account for the ICC in sample size calculations can result in an underpowered trial [4, 5, 11]. The second penalty identified

^aDepartment of Statistics, College of Math and Physical Sciences, The Ohio State University, Columbus, OH, USA

^bDivision of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, USA

^cDivision of Epidemiology, College of Public Health, The Ohio State University, Columbus, OH, USA

*Correspondence to: Michael L. Pennell, Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, USA.

†E-mail: mpennell@cph.osu.edu

by Cornfield was limited degrees of freedom; the number of groups randomized to each condition is usually small, which limits the degrees of freedom available for hypothesis testing and for estimating between-group variation. For example, Varnell, Murray, and Baker [12] pointed out that, in GRTs published between 1998 and 2002, 47% had fewer than 10 groups per condition. Thus, in a single study, there is usually little information about the between-group variation and, hence, the ICC. In addition to these two penalties, nowadays, studies collect data on multiple outcomes at several observation times. Such studies need a statistical model that can account for all these correlations: correlation caused by clustering (the ICC), correlation between different outcomes measured on the same individual, and over-time correlation (OTC) of each outcome while allowing treatment effects to change with time.

Our work is motivated by the Trial of Activity in Adolescent Girls (TAAG), which was a school-based GRT that had all the preceding characteristics. The goal of TAAG was to promote physical activity among middle school girls. In the study, three different cross-sectional samples were obtained from each participating school: a sample of sixth grade girls was drawn at baseline and samples of eighth grade girls were drawn 2 and 3 years after the intervention was introduced. At each time point, multiple outcomes were measured on consented girls including per cent body fat, activity level, BMI, and attitudes about PE class, as determined by questionnaire. A more detailed description of the design of the TAAG study is provided by Stevens *et al.* [1].

Most of the analyses of GRTs in the literature focus on single outcome models. Murray [4], Donner, and Klar [5], and Hayes and Moulton [11] provided comprehensive reviews of frequentist approaches, which use random effect models to account for ICC in cross-sectional and longitudinal analyses [4, 5, 11]. Bayesian hierarchical models have also been proposed for continuous [13] and binary [14, 15] outcomes, although these papers only address cross-sectional data. When multiple highly correlated outcomes are collected on each subject, joint modeling seems more sensible than fitting separate single outcome models. Studies in the field of reproductive and developmental toxicology have confirmed the advantages of joint modeling. Gueorguieva and Agresti [16] showed that joint modeling of multiple outcomes (birth weight and presence of malformations in rat pups) can provide significant efficiency gains when the outcomes are highly correlated and the sample size is not too large. General models for mixtures of discrete and continuous outcomes have been studied by Catalano and Ryan [17] and Rochon [18] from a generalized estimating equation perspective, by Dunson [19] from a Bayesian perspective, and by Regan and Catalano [20] and Gueorguieva and Agresti [16] from a maximum likelihood perspective. Very recently, Daniels and Normand [21] and Dunson [22] proposed correlated probit models for joint modeling of longitudinal data with multiple outcomes.

The joint models mentioned earlier are useful tools for analyzing complex cross-time multiple-outcome data structures; however, their priors do not reflect the nature of within-group correlations in community intervention trials and do not address Cornfield's second penalty: limited degrees of freedom. In particular, the Bayesian models in these papers utilized diffuse or noninformative priors on the variance components, which can lead to U-shaped priors placing high prior probability on small and large ICCs [13]. In GRTs, ICCs are usually small (e.g., ≤ 0.05) and, hence, a prior that places high probability on large values may be inappropriate. In recent years, researchers have been encouraged to publish ICC estimates from GRTs, and as a result, a substantial number and variety of ICCs have become publicly available [23]. It is thus desirable to incorporate ICC estimates from previous studies using Bayesian [24] or meta-analytic methods [25], thereby addressing the small degree of freedom problem of GRTs.

In this paper, we propose a Bayesian method for joint modeling continuous and binary outcomes collected at multiple time points in a GRT. Similar to Daniels and Normand [21], we model the outcomes through a latent multivariate normal regression model with shared random effects between the different outcomes. The model allows treatment effects to vary with time and accounts for three types of correlations: (i) ICC; (ii) the correlation between different outcomes measured on the same subject; and (iii) the OTC of each outcome. Moreover, we extend the method of Turner, Thompson, and Spiegelhalter [24] and propose an innovative prior structure for the variance components that directly places priors on the correlations; the priors reflect our knowledge about these correlations and allows the users to incorporate information from previous studies when they are available. In addition, our priors avoid undesirable constraints on the variance components that are imposed by some common methods and therefore yield unbiased estimates of the ICC, between outcome correlation, and OTC. We demonstrate through simulations the performance of our method under different ICCs and between outcome correlations. We also apply our method to obesity and body fat data from the TAAG study. To construct an informative prior for the ICC, we draw information from 10 previous studies on relevant outcomes in similar populations, and use a hierarchical approach to group these ICCs by outcome rather than by study.

Our paper is organized as follows: In Section 2, we describe the latent multivariate normal regression model for longitudinal multiple outcome data from GRTs. In Section 3, we describe our proposed prior structure, in particular, our new strategies for modeling the marginal variances in the latent normal model and for eliciting informative priors on the ICC based on information from previous studies. We next apply our method to simulated (Section 4) and TAAG data (Section 5), and we discuss the results and future work in Section 6.

2. Latent multilevel multidimensional normal regression model

We assume a GRT with a nested cross-sectional design as described by Murray [4]. Nested cross-sectional designs are used to examine the effect of an intervention on a large group of people. For example, the TAAG study used a nested cross-sectional design to determine if an educational program effectively increased physical activity levels within schools. In the basic nested cross-sectional design, independent random samples of subjects are drawn at baseline and follow-up. In this paper, we will assume that each subject is only sampled once (at baseline or follow-up), which is often the case.

Let there be two observation times ($t=1$ for baseline, $t=2$ for follow-up) and two outcomes collected for each individual at each time point. Without loss of generality, we assume that the first outcome ($k=1$) is continuous and the second outcome ($k=2$) is binary. Suppose that there are two treatment conditions (control versus intervention). Let there be G groups and assume that at time t ($t = 1, 2$), m_{it} members are randomly sampled from group i ($i = 1, \dots, G$). Let $y_{ijk t}$ be the k th outcome of the j th subject in group i measured at time t . We relate each outcome to an underlying normal random variable $z_{ijk t} \sim N(\mu_{ijk t}, \sigma_k^2)$ through a generalized link function: $y_{ijk t} = g_k(z_{ijk t})$ where $g_1(\cdot)$ is the identity link for the continuous outcome and, as in Albert and Chib [26],

$$g_2(z_{ij2t}) = I(z_{ij2t} > 0)$$

for the binary outcome, where $I(\cdot)$ is the indicator function.

Our model for the mean of each underlying normal random variable is the following:

$$\mu_{ijk t} = \beta_{0kt} + \beta_{1kt} \text{trt}_i + b_{ikt} + c_{ijt}, \tag{2.1}$$

where β_{0kt} is a fixed intercept for the k th outcome at time t , trt_i is an indicator variable that equals 1 if group i is assigned the intervention, and β_{1kt} is the fixed treatment effect at time t . Potentially, the model could be extended to include fixed effects of a set of covariates $x_{ijk t} = (x_{ijk t1}, \dots, x_{ijk tp})'$, which may vary with time and outcome. However, to simplify our presentation, we will assume no fixed effects in our model other than treatment. The remaining two terms in $\mu_{ijk t}$ are random effects. Similar to Daniels and Normand [21], we assume that the time-dependent group effects b_{ik1} and b_{ik2} follow a first-order Markov model, that is,

$$b_{ik1} \sim N(0, \sigma_{bk}^2),$$

and $b_{ik2} | b_{ik1} \sim N(\lambda_k b_{ik1}, (1 - \lambda_k^2) \sigma_{bk}^2),$ (2.2)

where σ_{bk}^2 measures between-group variability in the k th outcome and $\text{Corr}(b_{ik1}, b_{ik2}) = \lambda_k$. The conditional variance of $b_{ik2} | b_{ik1}$ was selected so that $\text{Var}(b_{ik1}) = \text{Var}(b_{ik2}) = \sigma_{bk}^2$. We also include random subject-specific effects (c_{ijt}) to accommodate correlations between different outcomes measured on the same subject and assume

$$c_{ijt} \sim N(0, \sigma_c^2). \tag{2.3}$$

This implies that the correlation coefficient between z_{ij1t} and z_{ij2t} (the two latent outcomes measured on the same subject at the same time) equals

$$\rho = \frac{\sigma_c^2}{\sqrt{(\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2)(\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2)}}. \tag{2.4}$$

3. Prior elicitation and model inference

Inferences are made from our model by using a Bayesian approach. Bayesian methods have many well-known advantages over frequentist methods including the ability to perform exact inferences using Markov chain Monte Carlo (MCMC) and the ability to incorporate information from previous studies through informative priors. The latter feature is especially pertinent to our research as we are interested in using published ICCs to address the small degrees of freedom problem in GRTs.

3.1. Prior for the fixed effects

We place a conventional diffuse prior on the fixed effects $(\beta_{0kt}, \beta_{1kt})'$, that is,

$$(\beta_{0kt}, \beta_{1kt})' \sim N_2(\mu_{\beta_{kt}}, \Sigma_{\beta_{kt}}) \quad (k = 1, 2), \quad (3.1)$$

where $\mu_{\beta_{kt}}$ is usually a vector of zeros and $\Sigma_{\beta_{kt}}$ is usually a diagonal matrix with large variances (e.g., both equal to 10,000). This prior is conjugate, so it greatly facilitates the posterior computation.

3.2. Priors for variance components

3.2.1. Priors for the residual error variance. We next consider modeling the variance of the latent z_{ijkt} conditional on the random effects, namely σ_1^2 and σ_2^2 . In generalized linear regression models, when $g_1(\cdot)$ is the identity link and $g_2(\cdot) = I(\cdot > 0)$, a common prior is

$$\sigma_1^{-2} \sim Ga(a_1/2, b_1/2),$$

where $Ga(a, b)$ denotes the Gamma distribution with mean a/b and variance a/b^2 , and σ_2^2 is fixed at 1 for identifiability considerations. This approach is perfectly valid when the two outcomes are modeled separately, because setting σ_2^2 to any positive constant σ_0^2 only rescales every term in the regression function for the binary outcome (μ_{ij2t}) by σ_0 . However, when the two outcomes are modeled jointly, this approach places undesirable constraints on the variance components. Suppose that we set σ_2^2 to a positive constant $\sigma_0^2 \neq 1$; consequently, the regression function for the binary outcome is rescaled to $\sigma_0 \mu_{ij2t}$ and the subject-specific random effect is rescaled to $\sigma_0 c_{ijt}$, which has variance $\sigma_0^2 \sigma_c^2$. Because c_{ijt} is present in both regression models (μ_{ij1t} and μ_{ij2t}), this rescaling also affects the model for the continuous outcome. However, z_{ij1t} , the variable underlying the continuous outcome is observed; thus, unlike the binary outcome, its variance cannot be arbitrarily rescaled, which results in biased estimates of the variance components. Under certain conditions, this prior also places a severe restriction on the between-outcome correlation (ρ). To see this, note that

$$\begin{aligned} \rho &= \frac{\sigma_c^2}{\sqrt{(\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2)(\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2)}} \\ &\leq \min \left(\frac{\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2}{\sqrt{(\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2)(\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2)}}, \frac{\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2}{\sqrt{(\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2)(\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2)}} \right) \\ &= \min \left(\frac{\sqrt{\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2}}{\sqrt{\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2}}, \frac{\sqrt{\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2}}{\sqrt{\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2}} \right). \end{aligned} \quad (3.2)$$

Therefore, when the two marginal variances $\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2$ and $\sigma_{b2}^2 + \sigma_c^2 + \sigma_2^2$ are very different, this prior specification restricts the between-outcome correlation ρ to be a small value. In situations where the outcomes are highly correlated, this restriction will make the estimate of ρ significantly biased.

To solve this problem, we propose a new method for specifying priors for the error variances. We first standardize the continuous outcome by using $z_{ij1t}^* = z_{ij1t}/std_1$, where std_1 is the sample standard deviation of the continuous variable. Assuming that z_{ij1t}^* satisfies the latent normal regression model (2.1), we then let

$$(\sigma_{b1}^2 + \sigma_c^2 + \sigma_1^2)^{-1} \sim Ga(a_{c1}/2, b_{c1}/2), \quad (3.3)$$

and fix $\sigma_{b_2}^2 + \sigma_c^2 + \sigma_2^2$ at 1. Because we standardize z_{ij1t} prior to analysis, $\sigma_{b_1}^2 + \sigma_c^2 + \sigma_1^2$ should be close to 1, and so is the upper bound of ρ . Thus, our new prior reduces the constraint imposed on the range of ρ and, in most cases, should allow it to be supported on nearly the full range $[0, 1]$. It also provides unbiased estimates of the variance components, which leads to unbiased estimates of the ICC.

3.2.2. *Priors for the ICC and the overtime correlation.* When analyzing GRTs, it is crucial to accurately estimate the ICC. Under our latent normal regression model (2.1), the ICC for outcome k at each time point is

$$\begin{aligned} \text{ICC}_k &= \text{Corr}(y_{ijkt}, y_{ij'kt}) \quad (j \neq j') \\ &= \frac{\sigma_{bk}^2}{\sigma_{bk}^2 + \sigma_c^2 + \sigma_k^2}. \end{aligned} \tag{3.4}$$

The traditional prior structure is to put independent inverse-gamma priors on σ_{bk}^2 , σ_c^2 , and σ_k^2 . As Spiegelhalter [13] pointed out, this structure implies a U-shaped prior on the ICC, with high prior probabilities assigned to ICCs close to zero and close to 1. However, in GRTs, the ICCs are typically small (e.g., $\text{ICC} \leq 0.05$) and, hence, priors that assign high weight to near perfect correlations do not make sense. Spiegelhalter proposed some more realistic priors for the ICC including a $Unif(0, 1)$ prior, a shrinkage prior, and an informative beta prior whose hyperparameters are estimated from data collected in previous studies.

More recently, Turner, Thompson, and Spiegelhalter [24] proposed more structured methods of constructing informative priors for the ICC. They weighted the ICC estimates from previous studies by their relevance to the outcome in the current study and then incorporated the information across several studies using a hierarchical structure. Turner *et al.*'s prior provides a very flexible structure for incorporating multiple relevant ICC estimates from several previous studies and increases the precision of the anticipated ICC. However, their model assumes that different outcomes from the same study are most related; we have found that among studies targeting the same type of group (e.g., school, worksite), the ICCs for the same outcome estimated in different studies tend to be more related than the ICCs for different outcomes from the same study. Also, in most cases, the relevance weights used in Turner *et al.*'s priors will be difficult to choose and, hence, are bound to be arbitrary. Therefore, we modify the Turner *et al.* approach to group the ICC estimates by outcome rather than by study. We only include the most relevant outcome from each study and assign equal weights to each ICC estimate.

Our development of the prior for the ICC begins by specifying a hierarchical model for the ICC estimates from previous studies. Let $\widehat{\text{ICC}}_{ks}$ denote the estimated ICC for outcome k ($k = 1, 2$) from the s th study ($s = 1, \dots, S_k$). Donner and Wells [27] showed that under large sample sizes

$$\widehat{\text{ICC}}_{ks} \sim N(\text{ICC}_{ks}, V(\text{ICC}_{ks}, M_s, G_s)), \tag{3.5}$$

where the variance, $V(\text{ICC}_{ks}, M_s, G_s)$, is approximated by Swiger's formula [28] using only the total number of subjects (M_s) and clusters (G_s), that is,

$$V(\text{ICC}_{ks}, M_s, G_s) = \frac{2(M_s - 1)(1 - \text{ICC}_{ks}^2)[1 + (M_s/G_s - 1)\text{ICC}_{ks}]}{(M_s/G_s)^2(M_s - G_s)(G_s - 1)}. \tag{3.6}$$

We then assign a normal prior to the logit of the true ICC from each study:

$$\text{logit}(\text{ICC}_{ks}) \sim N(\mu_{\text{ICC}_k}, \sigma_{\text{ICC}_k}^2) \tag{3.7}$$

where $\mu_{\text{ICC}_k} \sim N(m_{\mu k}, s_{\mu k}^2)$ and $\sigma_{\text{ICC}_k}^2 \sim Ga(a_{\text{ICC}_k}, b_{\text{ICC}_k})$. Finally, we assume that the ICC of outcome k in the current study (ICC_k) is drawn from the same distribution as $\text{ICC}_{k1}, \dots, \text{ICC}_{kS_k}$, which

results in the following conditional prior for ICC_k :

$$\begin{aligned} & \pi \left(\text{logit}(ICC_k) | \widehat{ICC}_k \right) \\ & \propto \int \int \int N(\text{logit}(ICC_k); \mu_{ICC_k}, \sigma_{ICC_k}^2) \\ & \quad \times \prod_{s=1}^{S_k} \left\{ N(\text{logit}(ICC_{ks}); \mu_{ICC_k}, \sigma_{ICC_k}^2) N(\widehat{ICC}_{ks}; ICC_{ks}, V(ICC_{ks}, M_s, G_s)) \right\} \\ & \quad \times N(\mu_{ICC_k}; m_{\mu k}, s_{\mu k}^2) Ga(\sigma_{ICC_k}^{-2}; a_{ICC_k}, b_{ICC_k}) \partial \mu_{ICC_k} \partial \sigma_{ICC_k}^{-2} \partial ICC_k^* \end{aligned} \quad (3.8)$$

where $\widehat{ICC}_k = (\widehat{ICC}_{k1}, \dots, \widehat{ICC}_{kS_k})'$, $ICC_k^* = (ICC_{k1}, \dots, ICC_{kS_k})'$, $N(x; m, v)$ denotes the pdf of the normal distribution with mean m and variance v evaluated at x , and $G(x; a, b)$ denotes the pdf of the gamma distribution with mean a/b evaluated at x . To implement our prior, the values of $m_{\mu k}$, $s_{\mu k}^2$, a_{ICC_k} , and b_{ICC_k} in the hyperpriors of μ_{ICC_k} and $\sigma_{ICC_k}^2$ must be chosen a priori. We recommend choosing a large value for $s_{\mu k}^2$ and small values for a_{ICC_k} and b_{ICC_k} , placing diffuse priors on μ_{ICC_k} and $\sigma_{ICC_k}^2$.

A similar method could be used to construct informative priors for the OTC. As noted earlier, TAAG used a nested cross-sectional design with an independent sample measured in each group at baseline and at follow-up. In this design, the OTC is the correlation between measurements from members from the same group measured at different time points. In terms of our model, the OTC for outcome k is

$$\begin{aligned} OTC_k &= \text{Corr}(y_{ijk1}, y_{ij'k2}) \\ &= \lambda_k \cdot \frac{\sigma_{bk}^2}{\sigma_{bk}^2 + \sigma_c^2 + \sigma_k^2} \quad (j = 1, \dots, m_{i1}; j' = 1, \dots, m_{i2}). \end{aligned} \quad (3.9)$$

Unfortunately, we are unaware of any estimates of group-specific OTCs and, hence, instead of using informative priors on the OTCs, we assign noninformative $Unif(0, 1)$ priors to λ_1 and λ_2 , the autocorrelations between the random group effects.

3.2.3. Prior for the between outcome correlation. Lastly, we put a noninformative prior on the between-outcome correlation, ρ , that is $\rho \sim Unif(0, 1)$. Of course, if previous data on ρ are available, one could replace the $Unif(0, 1)$ prior with an informative beta prior. Compared with the conventional method that places an inverse gamma prior on σ_c^2 , our prior is much easier to interpret and provides more direct estimates of the between outcome correlation.

3.2.4. Complete joint prior on the variance components. To summarize, the joint prior on our variance components is as follows:

$$\begin{aligned} \pi(\sigma_1^2 + \sigma_{b1}^2 + \sigma_1^2, ICC_1, ICC_2, \rho, \lambda_1, \lambda_2) &= Ga((\sigma_1^2 + \sigma_{b1}^2 + \sigma_1^2)^{-1}; a_{c1}/2, b_{c1}/2) \\ & \quad \times \pi(\text{logit}(ICC_1) | \widehat{ICC}_1) \pi(\text{logit}(ICC_2) | \widehat{ICC}_2) \\ & \quad \times I(0 \leq \rho, \lambda_1, \lambda_2 \leq 1), \end{aligned} \quad (3.10)$$

where $\pi(\text{logit}(ICC_k) | \widehat{ICC}_k)$ is our informative prior given in Equation 3.8, and $I(\cdot)$ denotes the indicator function. As mentioned in Section 3.2.1, $\sigma_2^2 + \sigma_{b2}^2 + \sigma_c^2$ is fixed at 1.

3.3. Methods for posterior inference

Under the preceding model specification, we design an MCMC algorithm to draw samples from the posterior distribution of the parameters. Our algorithm consists of both Gibbs and Metropolis–Hastings steps. The detailed conditional distributions may be found in Appendix A of the Supplementary Materials.[‡] Under mild regularity conditions, samples from this MCMC algorithm converge to a stationary distribution that is the joint posterior. We can then obtain parameter estimates using the posterior summaries of these samples. Sample R code for implementing our method is available upon request from the corresponding author.

[‡]Supporting information may be found in the online version of this article

4. Simulation study of potential efficiency gains

4.1. Methods

We performed a simulation study to investigate the efficiency gains provided by our method. We simulated data with the same structure as TAAG – nested cross-sectional design, two treatment groups, two time points (baseline and follow-up). The underlying normal random variables for the continuous and binary outcomes were generated from normal distributions with variance 1 and mean

$$\begin{aligned}\mu_{ijk1} &= -1 + trt_i + b_{ik1} + c_{ij1} \\ \mu_{ijk2} &= 2trt_i + b_{ik2} + c_{ij2},\end{aligned}$$

which means that the true effect of the intervention on the change in each outcome was 1. Random group effects (b_{ikt}) were generated assuming that λ_k (the autocorrelation between the group effects) was 0.5 (roughly equal to the posterior means when we applied our method to the TAAG data) and ICC values were generated from (3.7) with $\mu_{ICC_k} = \text{logit}(0.01)$ or $\text{logit}(0.1)$ and $\sigma_{ICC_k}^2 = 0.0001$. We also let ρ (between-outcome correlation) = 0.5 (medium correlation) or 0.8 (high correlation). Data were generated assuming six groups per treatment arm and 100 subjects per time point per group, and a total of 50 samples were generated for each of the four scenarios (two levels of ICC, two levels of ρ).

Each dataset was analyzed using three methods:

- (1) Latent multilevel multidimensional normal regression model (LMMNRM) with informative priors on the ICCs (our method).
- (2) LMMNRM with $Unif(0, 1)$ priors on the ICCs.
- (3) Separate latent multilevel normal regression models for each outcome with informative priors on the ICCs. Note that this method is synonymous with implementing Method 1 fixing all $c_{ijt} = 0$.

To construct informative priors on the ICC in Methods 1 and 3, we simulated ICCs from 10 previous studies (ICC_1, \dots, ICC_{10}) using (3.7) and their sample estimates ($\widehat{ICC}_1, \dots, \widehat{ICC}_{10}$) from (3.5) with each study consisting of 20 groups and 50 subjects per group. These sample estimates were then used to construct the informative priors for the ICC of the binary and continuous outcomes (Equation (3.8)). In each method, the following diffuse priors were specified for the hyperparameters μ_{ICC_k} and $\sigma_{ICC_k}^{-2}$:

$$\begin{aligned}\mu_{ICC_k} &\sim N(0, 10,000) \\ \sigma_{ICC_k}^{-2} &\sim Ga(0.001, 0.001).\end{aligned}$$

We ran our MCMC for a total of 80,000 iterations following a burn-in of 20,000 and saved every fifth value to thin the chain.

4.2. Results

Table I summarizes the results from our simulation study. The treatment effect referred to in the table is the difference in the pre–post change in the continuous and binary outcome across treatment arms, which is equivalent to the difference in the time-dependent intervention effects in our model, namely $\beta_{112} - \beta_{111}$ for the continuous outcome and $\beta_{122} - \beta_{121}$ for the binary outcome. In each simulation scenario, the biases of the treatment effect estimates were small and similar across methods. This result was expected; because of the identity link for the mean, the marginal effects of treatment provided by Method 3 are identical to the effects provided by Methods 1 and 2 conditional on c_{ijt} (the subject-specific effect). Also, because the historical ICC estimates used in our informative prior were drawn from a distribution centered on the true ICC (i.e., the prior is consistent with the data), we do not expect our informative prior to introduce bias in any of the parameter estimates.

Our method (Method 1) consistently provided the most efficient treatment effect estimates, although the effects on the binary outcome were more pronounced. The effects of joint modeling on efficiency were greatest when the two outcomes were strongly correlated; in Scenarios 1 and 3 ($\rho = 0.8$), the average posterior standard deviation of the treatment effect on the binary outcome was between 7–12% lower for our method compared with modeling the outcomes separately (Method 3) but only 2% lower in Scenarios 2 and 4 ($\rho = 0.5$). The finding that the benefits of joint modeling on efficiency (i) is greatest for the binary outcome; and (ii) increases with the between-outcome correlation is consistent with the results reported by Gueorguieva and Agresti [16].

Table I. Results from simulation study (estimates are averages across 50 datasets).

Scenario	Method ^b		Treatment effect ^a		ICC		ρ
			Continuous outcome	Binary outcome	Continuous outcome	Binary outcome	
1. ICC = 0.1 $\rho = 0.8$	1.	Mean	1.013	0.962	0.099	0.101	0.819
		S.D.	0.189	0.215	0.011	0.011	0.036
	2.	Mean	1.003	0.976	0.114	0.139	0.813
		S.D.	0.196	0.247	0.048	0.065	0.045
	3.	Mean	1.005	0.969	0.099	0.101	0
		S.D.	0.199	0.243	0.011	0.011	0
2. ICC = 0.1 $\rho = 0.5$	1.	Mean	1.023	1.027	0.101	0.098	0.542
		S.D.	0.191	0.242	0.011	0.011	0.048
	2.	Mean	1.024	1.036	0.129	0.115	0.581
		S.D.	0.203	0.253	0.057	0.058	0.074
	3.	Mean	1.022	1.025	0.101	0.099	0
		S.D.	0.192	0.247	0.011	0.011	0
3. ICC = 0.01 $\rho = 0.8$	1.	Mean	0.974	0.976	0.010	0.009	0.802
		S.D.	0.099	0.147	0.004	0.004	0.015
	2.	Mean	0.974	0.974	0.013	0.013	0.814
		S.D.	0.104	0.150	0.008	0.011	0.020
	3.	Mean	0.974	0.972	0.010	0.009	0
		S.D.	0.100	0.158	0.004	0.004	0
4. ICC = 0.01 $\rho = 0.5$	1.	Mean	0.990	1.013	0.009	0.009	0.494
		S.D.	0.098	0.158	0.004	0.004	0.025
	2.	Mean	0.990	1.013	0.013	0.017	0.499
		S.D.	0.102	0.166	0.009	0.014	0.026
	3.	Mean	0.990	1.017	0.009	0.009	0
		S.D.	0.098	0.160	0.004	0.005	0

^aEffect on change in outcome; i.e., $\beta_{1k2} - \beta_{1k1}$.

^b1. Joint model, informative prior on ICC; 2. Joint model, $ICC \sim Unif(0, 1)$; 3. Separate models, informative prior on ICC.

The informative prior on the ICC improved the efficiency of both the ICC and treatment effect estimates. In the case of the ICC, the improvement was dramatic. When Method 1 was implemented, the informative prior reduced the posterior standard deviation of the ICC of the binary outcome by approximately 80% in Scenarios 1 and 2 and approximately 70% in Scenarios 3 and 4 relative to Method 2, where a uniform prior was placed on the ICC. Similar results were observed for the ICC of the continuous outcome. The informative prior also improved the efficiencies of the treatment effect estimates, especially for the binary outcome. The greatest effect was observed in Scenario 1 ($ICC = 0.1, \rho = 0.8$) where our method exhibited a 13% reduction in the average posterior standard deviation of the treatment effect on the binary outcome relative to the joint model with a $Unif(0, 1)$ prior on the ICC (Method 2). In contrast, in Scenarios 2–4, the reduction in the posterior standard deviation of the effect on the binary outcome was only 2–5%. The differences in effect of the informative prior on efficiency of treatment effect estimates make sense when one thinks about the magnitude of the between group variance (σ_{bk}^2) relative to

$$\text{Var}(z_{ijkt}|c_{ijt}) = \sigma_{bk}^2 + \sigma_k^2;$$

the residual variance after accounting for the between-outcome correlation (ρ). Because our method fixes $\text{Var}(z_{ij2t}) = \sigma_{b_2}^2 + \sigma_c^2 + \sigma_2^2$ at 1 and rescales z_{ij1t} to $z_{ij1t}/\sqrt{\text{Var}(z_{ij1t})}$, we have $\sigma_{b_2}^2 = \text{ICC}_2$, $\sigma_c^2 = \rho$, and

$$\text{Var}(z_{ij2t}|c_{ijt}) = 1 - \rho.$$

In Scenario 1, the between group variability explains half of the residual variance and thus the informative prior results in a considerable reduction in the posterior standard deviation of the treatment effect. On the other hand, in Scenarios 2, 3, and 4, the between group variability only accounts for 20, 5, and 2% of the residual variance, respectively, and thus the informative prior has a negligible effect on the efficiency of the treatment effect.

As mentioned in Section 3.2.1, part of the novelty of our method is our priors for the residual error variances. The common practice of placing independent inverse gamma priors on the variance components imposes undesirable restrictions in joint models and may lead to biased estimates of the ICCs and the between-outcome correlation ρ . To illustrate this deficiency, we took the 50 datasets simulated under Scenario 3 ($\rho = 0.8$, $\text{ICC} = 0.01$), divided the continuous outcome by 10 (to make the variances of z_{ij1t} and z_{ij2t} different), and analyzed the data using a joint model with informative priors but set σ_2^2 to 1 and placed $\text{Ga}(0.001, 0.001)$ priors on σ_c^{-2} and σ_1^{-2} ; that is, the common approach used for latent normal regression models. Under these new priors, the treatment effect estimates remained the same (the average posterior means were within 3% of the true value), but the ICCs were considerably biased (average posterior means of ICC_1 and ICC_2 were 0.007 and 0.006, respectively) as was ρ (average posterior mean = 0.095). The bias of ρ is understandable because according to Equation 3.2, this method constrains ρ to be less than or equal to 0.1. As seen in the results for Methods 1 and 2 (Table I), our prior structure for the variance components successfully fixed this problem, yielding unbiased estimates for both the ICCs and ρ .

5. Analysis of the Trial of Activity in Adolescent Girls data

5.1. Description of the trial of activity in adolescent girls study and data

TAAG was a multicenter GRT, which examined a 2-year intervention targeting schools, community agencies, and girls intended to increase physical activity levels. A total of 36 schools were enrolled; 18 schools were randomly assigned to the intervention and 18 were assigned control. Random samples of sixth grade girls were drawn at baseline (average of 48 per school) and random samples of eighth grade girls were drawn 2 (average of 97 per school) and 3 years (average of 115 per school) later.

We illustrated our methodology through application to data collected at baseline and 3-year follow-up. We chose the 3-year follow-up instead of the 2-year follow-up for our analysis because the TAAG study found an effect of intervention on activity levels (the primary outcome) at the second follow-up but not the first [29]. Two groups (one per arm) were lost after the first follow-up leaving 34 groups for the analysis. The intervention effects of interest were those on change in average per cent body fat and obesity rate over this time interval. The latter outcome was measured by a binary variable, which equaled 1 if a girl's BMI was greater than the 95th percentile for her age and 0 otherwise. No covariates were included in our analysis.

5.2. Our models and prior elicitation

As in the simulation study, the TAAG data were analyzed using the following methods: (i) the LMMNRM with informative priors on the ICCs; (ii) the LMMNRM with $\text{Unif}(0, 1)$ priors on the ICCs; and (iii) separate latent multilevel normal regression models fit to each outcome. In Methods 1 and 3, informative priors for the ICCs were constructed using ICC estimates from previous studies. Ideally, the historical estimates used in constructing the prior should be from the same type of group as the application, but estimates from schools were unavailable for the outcomes we considered. Among the available ICC estimates, we selected those from groups similar to schools with respect to size and diversity of group members; this left us with estimates from 10 previous studies, among which three used worksite as the group of interest whereas the remaining seven reported clinic or physician as the group of interest. From each study, we chose the ICC corresponding to the outcome most relevant to obesity or per cent body fat, such as the ICCs for BMI, body fatness, weight, whether or not a subject had a BMI greater

than 30, and whether or not a subject was overweight. The ICC estimates and their references are provided in Table II. It turned out that, in each study, the ICC estimate most relevant to per cent body fat was also the ICC estimate most relevant to obesity and, hence, we used the same informative prior for the ICC of each outcome. The resulting prior distribution for the ICC was right-skewed with mean 0.03 and median 0.02 (Figure 1). The remaining priors were identical to those used in our simulations. For each method, we ran the MCMC for 100,000 iterations following a burn-in of 20,000, and we saved the results from every fifth iteration to thin the chain. We also repeated our analysis using a subset of 12 schools from California and Minnesota (six schools/state) to determine if improvements in efficiency were more evident in this smaller sample.

Reference	Outcome	Group	# Groups (G)	# Subjects (M)	ICC
Donner [30]	Body fatness ¹	Clinic/physician	20	5540	0.0130
Elley <i>et al.</i> [31]	BMI	Clinic/physician	42	861	0.0810
Jacobs <i>et al.</i> [32]	Weight	Worksite	7	2191	0.0177
Kelder <i>et al.</i> [33]	BMI	Worksite	32	5632	0.0190
Kinmonth <i>et al.</i> [34]	BMI	Clinic/physician	38	240	0.0450
Littenberg and MacLean [35]	BMI > 30	Clinic/physician	73	1044	0.0100
Martinson <i>et al.</i> [36]	Overweight ²	Worksite	24	10,067	0.0152
Parker <i>et al.</i> [37]	BMI	Clinic/physician	15	1175	0.0306
Smeeth and Ng [38]	BMI	Clinic/physician	53	13,785	0.0221
Turner <i>et al.</i> [24]	BMI	Clinic/physician	67	461	0.0210

¹ Ponderal index [height (in)/weight^{1/3} (lbs)] is below 12.5.

² BMI between 27.8 and 31.1 for men and BMI between 27.3 and 32.3 for women.

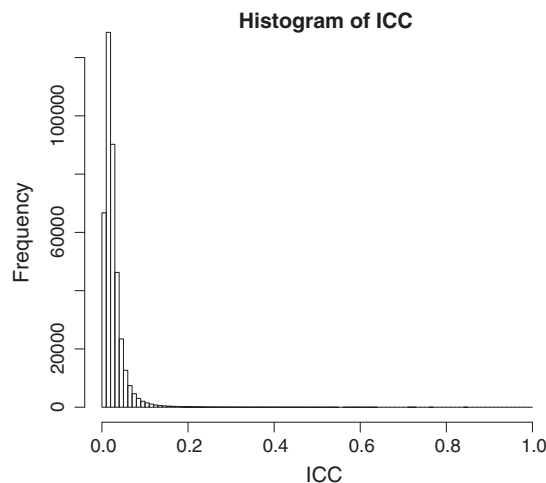


Figure 1. Informative prior on the ICC used in the TAAG analysis. Distribution was estimated by first drawing values of μ_{ICC_k} and $\sigma_{ICC_k}^2$ from their full conditional posteriors at each iteration of the MCMC and then drawing values of $logit(ICC_k)$ from a $N(\mu_{ICC_k}, \sigma_{ICC_k}^2)$ distribution.

Table III. Results from TAAG analysis.

Method ^b		Treatment effect ^a		ICC		ρ
		% Body fat	Obesity	% Body fat	Obesity	
1.	Mean	0.032	0.044	0.025	0.047	0.954
	S.D.	0.079	0.079	0.006	0.013	0.010
	95% C.I.	(-0.127, 0.177)	(-0.113, 0.204)	(0.015, 0.040)	(0.027, 0.077)	(0.928, 0.967)
2.	Mean	0.052	0.054	0.028	0.059	0.954
	S.D.	0.073	0.082	0.007	0.017	0.007
	95% C.I.	(-0.089, 0.213)	(-0.102, 0.227)	(0.016, 0.045)	(0.033, 0.100)	(0.938, 0.964)
3.	Mean	0.037	0.061	0.030	0.075	0
	S.D.	0.081	0.101	0.008	0.021	0
	95% C.I.	(-0.123, 0.198)	(-0.138, 0.261)	(0.018, 0.049)	(0.043, 0.125)	

^aEffect on change in outcome; i.e., $\beta_{1k2} - \beta_{1k1}$.

^b1. Joint model, informative prior on ICC; 2. Joint model, $ICC \sim Unif(0, 1)$; 3. Separate models, informative prior on ICC.

5.3. Results

Table III provides the posterior means, standard deviations, and 95% credible intervals of the treatment effects for Methods 1–3. As in the simulation study, the term treatment effect refers to the difference in pre–post change in each outcome across treatment arms. Whereas there were some differences in the posterior means across methods, the overall conclusions were the same; the intervention caused an increase in both average per cent body fat and obesity rate although the 95% credible intervals all contain zero, which is suggestive of an insignificant treatment effect. These results are consistent with those in Webber *et al.* [29]. They analyzed the complete TAAG dataset using an ANCOVA model applied to group-level means and found that per cent body fat and BMI were slightly higher in the intervention arm at the 3-year follow-up, although these differences were insignificant.

Although the conclusions about the intervention effect did not differ with method, their efficiencies differed considerably for the binary outcome. The posterior standard deviation for the treatment effect on obesity rate was smallest when the two outcomes were jointly modeled and when there was an informative prior on the ICC (Method 1) and largest when the outcomes were modeled separately (Method 3). For this dataset, joint modeling provided a greater benefit than informative priors: the posterior standard deviation of the treatment effect on obesity obtained from Method 1 (joint model, informative prior on ICC) was 22% smaller than the posterior standard deviation from Method 3 (separate models, informative prior on ICC) but only 4% smaller than the posterior standard deviation from Method 2 (joint model, uniform prior on ICC). This is because the correlation between outcomes was large (posterior mean of $\rho \approx 0.95$ as seen in Table I) and the informative prior had a relatively modest effect on the posterior standard deviation of the ICC of the binary variable (24% difference in the standard deviation between Methods 1 and 2). The modest impact of the informative prior on the standard deviation of the ICC was caused in part by the large sample size. Hence, to investigate the influence of sample size more thoroughly, we repeated our analysis using a subsample of 12 TAAG schools from California and Minnesota. As seen in Table B-1 in the Appendix, the posterior standard deviations of the ICC obtained from Method 1 were approximately half the posterior standard deviations from Method 2, that is, our informative priors on the ICC improved the estimation efficiency (defined in terms of the ratio of variances) by approximately 300%. As a result, the informative prior caused a greater reduction in the posterior standard deviation of the treatment effect on obesity; the posterior standard deviation under Method 1 was 7% lower than the value from Method 2.

Another interesting finding from our analysis of the TAAG data was that the ICC estimates varied with method. Whereas the posterior means of the per cent body fat ICC were similar across the three methods, the posterior mean of the obesity ICC was considerably higher when the outcomes were modeled separately. It is well known that regression adjustment for covariates can often reduce between-group variability, thereby decreasing the ICC [4]. Thus, it could be that the reason why joint modeling decreased the obesity ICC is that the subject-specific random effects (c_{ijt}) accounted for

missing covariates, which varied considerably by group. Joint modeling did not impact the ICC in the simulations because the only between group covariate was treatment arm, which was included in the model.

As in the simulation study, we also analyzed the TAAG data (34 groups) using a joint model with an informative prior on the ICC and the common approach of setting $\sigma_2^2 = 1$ and assigning diffuse inverse gamma priors on σ_c^2 and σ_1^2 . We did not standardize per cent body fat prior to this additional analysis. As in the simulation study, the treatment effect estimates (0.031 for per cent body fat and 0.055 for obesity) were similar to the estimates from Methods 1–3, but the ICC estimates (posterior mean = 0.001 and 0.002 for per cent body fat and obesity, respectively) were very different. However, unlike the simulation study, the posterior mean of ρ (0.948) was similar to the values in Methods 1 and 2; this was because the total variances of z_{ij1t} and z_{ij2t} were similar (posterior means = 76.235 and 72.923, respectively) and, hence, according to Equation 3.2, the upper bound of ρ was close to 1. It is interesting to note that even though the between outcome correlation was unaffected, the ICCs were deflated relative to Methods 1–3 because, as explained in Section 3.2.1, fixing σ_2^2 at 1 imposes undesirable restrictions on the variance components of both the continuous and binary outcome.

6. Discussion

In this paper, we developed a joint model for binary and continuous outcomes collected in GRTs. We have also extended previous work by Turner *et al.* [24] to place an informative prior on the ICC of each outcome, which, in combination with joint modeling, improves efficiency, particularly in the case of the binary outcome. Whereas the work presented was limited by one data example and a simulation study consisting of a limited number of scenarios, the gains in efficiency that we observed are consistent with previous work. In a simulation study, Gueorguieva and Agresti [16] reported that joint modeling resulted in a 17% reduction in the standard error of the treatment effect on the binary outcome when the true correlation between the continuous and binary outcome was 0.8; similarly, we found that when $\rho = 0.8$ in our simulations, joint modeling caused a 7–12% reduction in the posterior standard deviation of the effect of treatment on the binary outcome. In an analysis of data from a previous GRT, Turner *et al.* [24] found that placing an informative prior on the ICC caused a 10% reduction in the posterior standard deviation of the intervention effect; in our simulations, we found that the informative prior reduced the average posterior standard deviation of the treatment effect on the binary outcome by a maximum of 13% across the scenarios considered. Whereas a 12–13% reduction in the standard deviation may sound modest, it could be very beneficial in a GRT. For example, when designing a study, if we know that we will be using an analysis method that decreases the standard deviation of the treatment effect by 12% relative to a univariate mixed model, this could mean a reduction in sample size of one group per arm, which would significantly reduce study costs in many GRTs, particularly those targeting large groups of people (e.g., counties).

Our simulation studies demonstrated the behavior of our method under different levels of ICC and between outcome correlation. As in Gueorguieva and Agresti [16], we found that joint modeling causes the greatest improvement in efficiency when the two outcomes are strongly correlated. In the case of the binary outcome, we found that the informative prior causes the greatest improvement in efficiency when a substantial proportion of the residual variance is caused by between-group variability (or ICC); to the best of our knowledge, we are the first to present such a finding. Thus, our recommendations for future use are as follows: when the between-outcome correlation between a continuous and binary outcome is strong, we recommend joint modeling. If the ICC is large (particularly for the binary variable), we recommend extending the approach to include an informative prior on the ICC, otherwise, a uniform prior on the ICC is probably sufficient if all one cares about is improving the efficiency of the treatment effect estimates. Weakly correlated outcomes could be analyzed using our method, but the results would probably be similar to univariate models for each outcome, and, hence, the benefit is probably too small relative to the added computational complexity.

Evaluation of model fit is an important part of any modeling exercise. Unfortunately, we are unaware of a formal method for evaluating the fit of a joint model for a binary and continuous response. Future research should consider developing a method for measuring goodness of fit of our model, which allows direct comparison to competing models including univariate models for each outcome and models with noninformative priors.

In this paper, we focused our efforts on nested cross-sectional GRTs. Another design commonly used is the nested cohort design where members in each group are followed over time. Potentially, our methodology could be extended to handle these designs with the major modification being that the model for the mean must include random subject-specific effects for each outcome in addition to shared random effects across outcomes. We also focused on joint modeling of binary and continuous outcomes; whereas these are probably the two most common types of outcomes in GRTs, ordinal, count, and survival outcomes are also collected and, hence, joint modeling of these outcomes presents an interesting area of future research.

Finally, our model examined the effects of treatment on change in each outcome (post–pre), which is a common approach in GRTs, but not the only option. For example, Webber *et al.* [29] analyzed the TAAG data using a two-stage approach; in the first stage, a mixed model was used to regress each outcome on fixed time and ethnicity effects and random school and school-by-time interaction effects. This model was used to generate ethnically adjusted pre-test and post-test means for each group, which were inserted into a second-stage ANCOVA model in which the adjusted post-test mean was regressed on the adjusted pre-test mean and treatment group. Future research should examine the potential of joint modeling and informative priors within two-stage and other group-level analyses.

Acknowledgements

The project described was supported by Award Number UL1RR025755 from the National Center for Research Resources (Xu, Pennell, and Lu) and Award Number DMS-09-07070 from the National Science Foundation (Xu). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Research Resources or the National Institutes of Health. TAAG was funded by the following cooperative agreements from the National Heart, Lung and Blood Institute: U01 HL066855 (Tulane University); U01HL066845 (University of Minnesota); U01HL066852 (University of South Carolina); U01HL066853 (University of North Carolina at Chapel Hill); U01HL066856 (San Diego State University); U01HL066857 (University of Maryland); U01HL066858 (University of Arizona).

References

1. Stevens J, Murray DM, Catellier DJ, Hannan PJ, Lytele LA, Elder JP, Young DR, Simons-Morton DG, Webber LS. Design of the trial of activity in adolescent girls (TAAG). *Contemporary Clinical Trials* 2005; **26**:223–233.
2. Beresford SAA, Locke E, Bishop S, West B, McGregor BA, Bruemmer B, Duncan GE, Thompson B. Worksite study promoting activity and changes in eating (PACE): design and baseline results. *Obesity* 2007; **15**:4S–15S.
3. Paskett ED, Tatum CM, D'Agostino RB, Rushing J, Velez R, Michielutte R, Dignan M. Community-based interventions to improve breast and cervical cancer screening: results of the Forsyth County Cancer Screening (FoCaS) Project. *Cancer Epidemiology, Biomarkers and Prevention* 1999; **8**:453–459.
4. Murray DM. *Design and Analysis of Group-Randomized Trials*. Oxford University Press: New York, 1998.
5. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Oxford University Press: New York, 2000.
6. Feder G, Griffiths C, Eldridge S, Spence M. Effect of postal prompts to patients and general practitioners on the quality of primary care after a coronary event (POST): randomised controlled trial. *British Medical Journal* 1999; **318**:1522–1526.
7. Cornfield J. Randomization by group: a formal analysis. *American Journal of Epidemiology* 1978; **108**:100–102.
8. Murray DM, Hannan PJ, Baker WL. A Monte Carlo study of alternative responses to intraclass correlation in community trials: is it ever possible to avoid Cornfield's penalties? *Evaluation Review* 1996; **20**:313–337.
9. Murray DM, Wolfinger RD. Analysis issues in the evaluation of community trials: progress toward solutions in SAS/STAT MIXED. *Journal of Community Psychology* 1994; **CSAP Special Issue**:140–154.
10. Zucker DM. An analysis of variance pitfall: the fixed effects analysis in a nested design. *Educational and Psychological Measurement* 1990; **50**:731–738.
11. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. CRC Press: Boca Raton, FL, 2009.
12. Varnell SP, Murray DM, Baker WL. An evaluation of analysis options for the one group per condition design: can any of the alternatives overcome the problems inherent in this design? *Evaluation Review* 2001; **25**:440–453.
13. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; **20**:435–452.
14. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**:453–472.
15. Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine* 2007; **26**:2088–2112.
16. Gueorguieva RV, Agresti A. A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* 2001; **96**:1102–1112.
17. Catalano PJ, Ryab LM. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* 1992; **87**:651–658.

18. Rochon J. Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics* 1996; **52**:740–750.
19. Dunson DB. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* 2000; **62**:355–366.
20. Regan MM, Catalano PJ. Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. *Biometrics* 1999; **55**:760–768.
21. Daniels MJ, Normand ST. Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics* 2006; **7**:1–15.
22. Dunson DB. Bayesian methods for latent trait modelling of longitudinal data. *Statistical Methods in Medical Research* 2007; **16**:399–415.
23. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health* 2004; **94**:423–432.
24. Turner RM, Thompson SG, Spiegelhalter DJ. Prior distributions for the intraclass correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. *Clinical Trials* 2005; **2**:108–118.
25. Blitstein JL, Hannan PJ, Murray DM, Shadish WR. Increasing the degrees of freedom in existing group randomized trials: the df_* approach. *Evaluation Review* 2005; **29**:241–267.
26. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*; **88**:669–679.
27. Donner A, Wells G. A comparison of confidence interval methods for the intraclass correlation coefficient. *Biometrics* 1986; **42**:401–412.
28. Swiger LA, Harvey WR, Everson DO, Gregory KE. The variance of intraclass correlation involving groups with one observation. *Biometrics* 1964; **20**:818–826.
29. Webber LS, Catellier DJ, Lytle LA, Murray DM, Pratt CA, Young DR, Elder JP, Lohman TG, Stevens J, Jobe JB, Pate RR. Promoting physical activity in middle school girls: trial of activity for adolescent girls. *American Journal of Preventive Medicine* 2008; **34**:173–184.
30. Donner A. An empirical study of cluster randomization. *International Journal of Epidemiology* 1982; **11**:283–286.
31. Elley CR, Kerse N, Chondros P, Robinson E. Intraclass correlation coefficients from three cluster randomized controlled trials in primary and residential health care. *Australian and New Zealand Journal of Public Health* 2005; **29**:461–467.
32. Jacobs DR, Jeffrey RW, Hannan PJ. Computation of variance in worksite data: unit of analysis. In *Methodological Issues in Worksite Research: Proceedings of a Workshop*. Johnson K, LaRosa JH, Scheirer CJ, Wolle JM (eds). Airlie House, VA: National Heart, Lung, and Blood Institute, NIH, 1982; 77–88.
33. Kelder SH, Jacobs DR, Jeffrey RW, McGovern PG, Forster JL. The worksite component of variance: design effects and the healthy worker project. *Health Education Research* 1993; **8**:555–566.
34. Kinmonth AL, Woodcock A, Griffin S, Spiegel N, Campbell MJ. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. *BMJ* 1998; **317**:1202–1208.
35. Littenberg B, MacLean CD. Intra-cluster correlation coefficients in adults with diabetes in primary care practices: the Vermont Diabetes Information System field study. *BMC Medical Research Methodology* 2006; **6**:20. DOI: 10.1186/1471-2288-6-20.
36. Martinson BC, Murray DM, Jefferey RW, Hennrikus DJ. Intraclass correlation for measures from a worksite health promotion study: estimates, correlates and applications. *American Journal of Health Promotion* 1999; **13**:347–357.
37. Parker DR, Evangelou E, Eaton CB. Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART). *Contemporary Clinical Trials* 2005; **26**:260–267.
38. Smeeth L, Ng E. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC trial of the assessment and management of older people in the community. *Controlled Clinical Trials* 2002; **23**:409–421.