A Bayesian model for the common effects of multiple predictors on mixed outcomes

Robert E. Weiss¹, Juan Jia¹,² and Marc A. Suchard¹,³,⁴,*

¹Department of Biostatistics, UCLA School of Public Health, University of California, Los Angeles, CA 90095-1772, USA
²Amgen Inc., South San Francisco, CA 94080, USA
³Departments of Biomathematics, and ⁴Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA

We propose a Bayesian multivariate model in which a single linear combination of the covariates predict multiple outcomes simultaneously. The single linear combination is a data-derived score along the lines of the Apache or Charlson index scores for critically ill patients, the Karnofsky or Eastern Cooperative Oncology Group score for cancer patients or Euro-score for cardiac patients that may be used to predict multiple outcomes. Outcomes may be discrete or continuous and we use a composition of generalized linear models for the marginal distribution for each outcome. We explain how to set the prior distribution and we use Markov chain Monte Carlo methods to calculate the posterior distribution. We propose two types of expanded models to diagnose whether each outcome indeed has predictor effects common with the other outcomes, and whether a particular predictor is commonly predictive for all outcomes. We determine a final model based on the diagnostic models. The method is applied to a study yielding multiple psychometric outcomes of mixed type measured in young people living with human immunodeficiency virus.

Keywords: Bayesian Wald test; human immunodeficiency virus; index construction; multivariate regression; single index model

1. INTRODUCTION

Multiple outcomes are common in medical, psychological and sociological studies. Joint analysis of continuous multiple outcomes is common in recent decades [1–18]. Difficulty arises when the outcomes comprise a mix of continuous, binary and non-negative integer variables. In recent years, a number of models have been developed to handle mixed outcomes, e.g. [19–24].

Little & Schluchter [19] and Gueorguieva & Agresti [24] have proposed maximum-likelihood procedures for analysing mixed continuous and binary data with missing values. Little & Schluchter model q categorical variables defining a q-way contingency table with $C = \prod I_i$ cells where the $i$th categorical variable has $I_i$ levels. They assumed a multinomial distribution over the $C$ cells of the contingency table. Given the contingency table cells, the continuous outcomes follow a multivariate normal distribution. Sparse data can make inference problematic. Gueorguieva & Agresti proposed an underlying multivariate probit model for the binary variables that correlate directly with the continuous normal variables.

Sammel et al. [20] proposed a latent variable model for mixed discrete and continuous outcomes with the multiple outcomes correlated through subject-specific latent variables. Conditional on the latent variables, outcomes were assumed independent with each outcome arising from a one parameter exponential family, with its mean a function of the latent variable and other covariates. An expectation-maximization (EM) algorithm was used to fit the model.

Arminger & Kusters [21] assumed each outcome had a latent variable and jointly, the latent variables follow a multivariate normal distribution with expectation potentially dependent on covariates. Dunson [23] proposed a Bayesian latent variable model for clustered mixed outcomes. It generalized Arminger & Kusters’ work to accommodate non-normal latent variables for multi-level data. It allowed nonlinear relationships between the covariates and latent variables and used multiple latent variables for each type of outcome and covariate-dependent modification of the relationship between the latent and covariates.

In this paper, we propose a double link common predictor effect (COPE) model in which a single linear combination of the covariates is used to predict all outcomes simultaneously. Generalized linear models are used to model the marginal distribution for each outcome.
Outcome-specific double link functions connect each outcome’s mean to the single linear predictor, where a double link refers to the composition of one link function in a second link function. Correlations among outcomes are modelled by latent variables. Markov chain Monte Carlo (MCMC) algorithms are used to estimate the posterior distribution of the parameters and the latent variables.

Compared with the model proposed by Sammel et al. [20] in which each outcome has an outcome-specific set of fixed effect coefficients, our model furnishes a parsimonious common set of fixed effects for all outcomes, thus we have greater efficiency in our parameter estimates. Compared with Dunson’s model, in which all outcomes have a common set of fixed effects, our model returns greater flexibility for modelling COPE through the double link function between the linear predictor and outcomes. In general, the usual multiple outcome model is the saturated model, which has a separate set of coefficients for each outcome. Our model, when appropriate, is much more efficient as there are many fewer coefficients.

Frequentist methods are based on asymptotic maximum-likelihood theory and the standard errors and confidence interval are based on asymptotic normality assumptions. In Bayesian inference, posterior distribution of parameters and latent variables can be estimated by MCMC methods, and means and variances for parameters can be appropriately estimated without requiring a large sample size. For our proposed model, a key inference involves the product of parameters; using a maximum-likelihood approach requires the delta method to estimate point estimates and variances of the product of parameters. With the Bayesian approach, inferences can be drawn directly from posterior samples of the product of the parameters.

Index scores for evaluating disease severity are common in medicine. Examples include the Apache [25,26] and Charlson index scores [27,28] for critically ill patients, the Karnofsky and Eastern Cooperative Oncology Group scores for cancer patients [29,30] and EuroScore [31,32] for cardiac patients. These scores are often used for predicting multiple outcomes beyond the outcome used for the initial derivation. The single linear combination estimated by our model is a principled data-derived score along the lines of these scores that may be used to predict multiple outcomes. Our Bayesian methodology using MCMC methodology makes it easy to produce both (i) index calculations for an individual complete with uncertainty intervals and (ii) predictions for single or multiple outcomes. Construction of online calculators is straightforward using MCMC methodology.

This paper is organized as following: §2 proposes the double link COPE model for mixed outcomes. Section 3 explains the prior specification. In §4, we relax our insistence on only one linear combination of the covariates for predicting all outcomes, and we describe two types of extended models both as model diagnostics and as potentially valuable models in their own right. Section 5 applies the method to multiple, mixed-type psychometric outcomes measured on young people living with human immunodeficiency virus (YPLH) and the paper concludes with a discussion in §6.

2. COMMON PREDICTOR EFFECT MODEL SPECIFICATION

We observe $k_i$ continuous outcomes, $k_2$ binary outcomes, $k_3$ count outcomes and $L$ covariates on subject $i = 1, \ldots, n$. Let $y_{ik}$ denote the $k$th outcome for subject $i$, $k = 1, \ldots, K$, $K = k_1 + k_2 + k_3$, and define the $i$th observation $y_i = (y_{i1}, \ldots, y_{ik})$; let $x_{id}$ denote the $d$th predictor, $d = 1, \ldots, L$, for subject $i$ with $x_{id} = (x_{i1d}, \ldots, x_{id})$, where $x_{id}$ does not include an intercept. Covariates are the same for all outcomes and a single linear predictor, meaning a single linear combination of covariates, $x_i' \alpha$, will be used to predict all outcomes.

An outcome-specific double link function is needed for each outcome in our proposed model: the usual link function $g_k(\cdot)$ for a generalized linear model that connects the expected value of subject $i$ outcome $k, \mu_k$, to the natural parameter $\theta_{ik}$: and a location-scale link function $\psi_k(\cdot)$ that connects its natural parameter with the single linear predictor. In our model, we take $\psi_k(\cdot)$ to be linear, $\psi_k(x_i'\alpha) = \gamma_0k + \gamma_{1k}x_i'\alpha$, which allows the COPE for each outcome to be modified by a scale parameter. The usual link functions for generalized linear models are the identity link for continuous outcomes $\eta_{ik} = g_k(\theta_{ik})$; the logit link for binary outcomes $\eta_{ik} = g_k(\theta_{ik}) = \log(\theta_{ik}/(1-\theta_{ik}))$, where $\log(x) = \log(x - 1 - x^{-1})$ and its inverse, called expit, $\text{expit}(x) = \exp(x)/(1 + \exp(x))^{-1}$; and the log link for count outcomes $\eta_{ik} = g_k(\theta_{ik}) = \log(\theta_{ik})$. We model the correlations across the multivariate outcomes on the linear predictor scale through a random effect variable $\beta_{ik}$, $\eta_{ik} = \psi_k(x_i'\alpha) + \beta_{ik}$.

The double link COPE model for mixed outcomes is

\[\eta_{ik} = \psi_k(x_i'\alpha) + \beta_{ik} = \gamma_0k + \gamma_{1k}x_i'\alpha + \beta_{ik}. \tag{2.1}\]

\[y_{ik}|\beta_{ik} \sim N(\eta_{ik}, \sigma^2), \quad k = 1, \ldots, k_3 \tag{2.2a}\]

\[y_{ik}|\beta_{ik} \sim \text{Bernoulli}(\expit(\eta_{ik})), \quad k = k_1 + 1, \ldots, k_1 + k_2 \tag{2.2b}\]

\[y_{ik}|\beta_{ik} \sim \text{Poisson}(\exp(\eta_{ik})), \quad k = k_1 + k_2 + 1, \ldots, K, \tag{2.2c}\]

where $\alpha$ is an $L \times 1$ vector of fixed-effect regression coefficients for all outcomes, $\gamma_0k$ and $\gamma_{1k}$ are the intercept and scale parameter in the link for outcome $k$, respectively, $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0K})'$, $\gamma_1 = (\gamma_{11}, \ldots, \gamma_{1K})'$, $\beta = (\beta_1, \ldots, \beta_{ik})'$ follows a multivariate normal distribution with mean 0 and covariance $\Sigma(\xi)$, $\beta \sim N_K(0, \Sigma(\xi))$, where $\Sigma(\xi)$ models the correlations among the multiple outcomes on the linear predictor scale, and $\xi$ is a vector of unknown parameters for the components of the correlation matrix.

As a whole, $\gamma_{1k}x_i'\alpha$ can be estimated in model (2.1); however, there is an identifiability problem if we want to estimate $\alpha$ and $\gamma_i$ individually. For example, $\alpha$ and $\gamma_i$ can be replaced by $\tilde{\alpha} = a\alpha + c$ and $\gamma_{ik} = \gamma_{ik}/c$ for an arbitrary constant $c \neq 0$ and we still have the same prediction. We need to constrain $\alpha$ and/or $\gamma_i$ to solve the problem. The solution we propose is to divide the regression coefficients by their Euclidean norm $|\alpha| = (\alpha'\alpha)^{1/2}$ and fix the sign of the regression coefficient for
one pre-specified predictor whose regression coefficient is significantly different from 0, without lack of generality, say \( \gamma_k \). We modify \( \eta_k \) in model (2.1) as
\[
\eta_k = \psi_k(x_i' \alpha, \beta_k) = \gamma_k + \gamma_k x_i' \alpha + \beta_k. 
\]

3. PRIOR SPECIFICATION

We specify prior distributions for the parameters in our proposed model to complete the Bayesian modelling. Many researchers have discussed different ways to specify priors [33–36]. Prior information may come from previous studies, published results from similar studies, scientific reasoning about the measured variables, other related data or expert opinions. For any model, it is helpful to have an automatic prior that requires relatively little in the way of substantive input to allow data and model exploration to begin. Our own preference is to add scientific/statistical reasoning on top of that to allow for sensible informative proper priors particularly for parameters not of primary scientific interest. In contrast, specifying a fully subjective prior requires a large investment in time and training of subject matter specialists and is not an activity that can be lightly undertaken.

For models (2.2a)–(2.3), we specify the prior distributions for common regression coefficients \( \alpha \), intercepts \( \gamma_0 \), scale parameters \( \gamma_k \) and the covariance \( \Sigma \) of the latent variables \( \beta_k \). We first standardize each continuous predictor by subtracting its mean and dividing by its standard deviation, the resulting continuous predictor has mean 0 and standard deviation 1. We assign a normal prior to \( \alpha, \beta_k \sim N(\mu_\alpha, \Sigma_\alpha) \), where \( \mu_\alpha \) is a vector of location parameters and \( \Sigma_\alpha \) is a covariance matrix. We set \( \mu_\alpha = 0L \), an \( L \)-vector of 0s, and \( \alpha \)'s to be independent, \( \Sigma_\alpha = \sigma^2_a I_L, I_L \) is the \( L \times L \) identity matrix, and we give \( \sigma^2_a \) an arbitrary value \( c \). This is a flat prior for \( || \alpha || \) on the unit ball in \( \mathbb{R}^L \). The sampling distribution of \( Y \) is not affected by the length \( || \alpha || \) of \( \alpha \) and the posterior of \( || \alpha || \) is equal to its prior.

We discuss priors for the intercept \( \gamma_0 \) and scale parameters \( \gamma_k \) separately for each type of outcome: Bernoulli, count and continuous. For Bernoulli outcomes, a wide range of probabilities of a response equal to 1 is from 0.001 to 0.999 (0.001 < \( P(\gamma_0 = 1) < 0.999 \)), which gives a range after logit transformation of \(-6.9 \) to 6.9. We assign a normal prior to \( \gamma_0 \) with mean 0 and variance \((6.9 - (-6.9))^2 = 11.9 \) for binary outcomes, which is the square of the range divided by 4, inspired by the idea that in the prior, the mean \pm 2 s.d. should cover 95 per cent of the prior range. For count outcomes, we take the mean of the response to be somewhere between 0.01 and 100. After a log transformation, we have a range of the linear predictor of \(-4.6 \) to 4.6. We assign a normal prior to \( \gamma_0 \) with mean 0 and variance \((4.6 - (-4.6))^2 = 53 \) for count outcomes. For continuous outcomes, we assign a normal prior to \( \gamma_0 \) with mean \( \gamma_k \) and variance \((\gamma_k^2 / 2)^2 \). We consider these priors to be vague and not particularly informative although proper. The prior for continuous outcome uses the data, but the information in the prior is limited. Another approach for continuous outcomes is to guess at the range \( = (\max - \min) \) of each outcome and use the mid-range, \((\max + \min) / 2 \), as prior mean and \((\max - \min) / 4 \) as prior variance.

In equation (2.3), scale parameter \( \gamma_k \) is the slope of the regression of \( \eta_k \) on \( x_i / || \alpha || \). When the linear predictor is perfectly predictive, the absolute value of the slope can be estimated as the range \( (\max - \min) \) of \( \eta_k \) divided by the range \( (\max - \min) \) of \( x_i / || \alpha || \), with an appropriate sign either positive or negative. When the linear predictor is not predictive at all, \( \gamma_k = 0 \). We assign a normal prior with mean 0 and variance the square of the range of the natural parameter divided by 2 times the range of single linear predictor
\[
\frac{\max(\eta_k) - \min(\eta_k)}{2(\max(|x_i / || \alpha |||) - \min(|x_i / || \alpha |||))}^2. 
\]

We previously discussed the range of the natural parameters for binary, Poisson and continuous outcomes when specifying priors for the intercept. We now discuss the range of \( x_i / || \alpha || \). We deal with continuous predictors and binary predictors separately. The standardized continuous predictors, the norm of the fixed effect regression coefficients equal to 1 implies that the absolute value of each regression coefficient \( a_i / || \alpha || \) is less than 1; an approximate range of \( x_i a_i / || \alpha || \) is then at most roughly \(-2 \) to \( 2 \). For binary predictors, we have
\[
\sum_{l=1}^L x_l a_l \leq \sqrt{\sum_{l=1}^L a_l / || \alpha ||} \leq \sqrt{1 / a_0} \leq \sqrt{L}. 
\]

Here \( l \) indexes the binary predictors, \( L \) is the total number of binary predictors, \( 1 \) is a vector of 1 with length \( L \) and \( a_0 / || \alpha || \) is the vector of regression coefficients for the binary predictors. The range of \( x_i a_i / || \alpha \) is therefore roughly \( 4*|L - L| + 2*\sqrt{L} \), where \( L - L \) is the number of continuous predictors. We substitute this range in place of the denominator of equation (3.1) to calculate the prior variances for \( \gamma_k \).

We use a Wishart prior \( \Sigma^{-1} \sim \text{Wishart}(v, A) \), for precision matrix \( \Sigma^{-1} \), with degrees of freedom \( v \) and precision \( A \), where \( \Sigma^{-1} \) is the inverse of covariance matrix \( \Sigma \) in model (2.3). We set \( v \) to be a number somewhat bigger than \( K \), the dimension of \( A \), and we set \( A \) to be a diagonal matrix. We set diagonal element \( A_{kk} \) to \( v - K - 1 \) divided by a prior estimate of the variance of \( \eta_k \). For continuous outcomes, we use the prior-estimated range divided by 4 as our prior estimate of the variance of \( \eta_k \) for binary outcomes, we use \((6.9 - (-6.9))/4)^2 = 11.9 \) and for count outcomes, we use \((4.6 - (-4.6))/4)^2 = 5.3. \]

4. DIAGNOSTIC MODELS

A priori, outcomes are chosen to share a single linear predictor. In practice, we may choose outcomes which we hope are appropriate. Similarly, we may have chosen predictors that are generally predictive of all outcomes, however some of them may not belong to a single linear predictor. The COPE model assumes that all outcomes are predicted by the single linear predictor and all predictors are commonly predictive of all outcomes. There are models whose number of fixed effects parameters is in between the COPE model and the saturated model, and we propose two extended models to
check the COPE model assumptions. Model A checks whether one pre-specified outcome belongs with other outcomes, and model B checks whether one pre-specified predictor is commonly predictive. We apply model A to all outcomes in turn. Any outcomes that do not belong to the COPE model are removed and are fit separately barring some value to continuing to fit a multivariate model. We then apply model B to all predictors in turn. In contrast to outcomes, predictors often can not be arbitrarily dropped for statistical reasons and must be kept in the model for scientific reasons. Thus, we may actually prefer the model B version of the COPE model—a kind of relaxed COPE model where most covariates belong to a single linear predictor, but a few covariates are not required to be part of the COPE linear predictor.

4.1. Extended model A: one outcome has a separate set of predictor effects

For a pre-specified outcome s, we modify the linear predictor in model (2.3) by adding an extra set of regression coefficients for outcome s on top of the common regression coefficients, to allow the fixed effects for outcome s to differ from the common effects while keeping the linear predictor for other outcomes the same as in model (2.3)

\[
\eta_{s} = \gamma_{0s} + \gamma_{1s}x_{ik}\frac{\alpha}{\|\alpha\|} + x_{i}s\alpha_{s},
\]

where \(\alpha_{s}\) is the extra set of regression coefficients, \(\alpha_{s}\) is the departure from the common effect for all outcomes of the effect of predictor \(l\) on outcome \(s\), \(l = 1, \ldots, L\). We assign \(\alpha_{s}\) a normal prior with mean 0 and variance \(I_{L}\).

Models (2.2a)–(4.1) have an identifiability problem, \(\gamma_{s}\) and \(\alpha_{s}\) can be replaced by \(\gamma_{s} = \gamma_{1s} - c\) and \(\alpha_{s} = \alpha_{s} + c\frac{\alpha}{\|\alpha\|}\) for an arbitrary constant \(c\). To solve the problem, we use the Gram–Schmidt process to split \(\alpha_{s}\) into two components, one component \(c^*\) is parallel to \(\alpha_{s}\), where \(c^* = \alpha_{1s}^{*}/\alpha_{s}\) and the other component, denoted \(\alpha_{s}^{*}\), \(\alpha_{s} = -c^*\alpha_{s}\), is perpendicular to \(\alpha_{s}\), \(\alpha_{s}^{*}\). We use MCMC sampling for computations and we calculate \(\alpha_{s}^{*}\) for each posterior sample of \(\alpha_{s}\).

Having fit the model, we test the hypothesis \(H_{0s}: \alpha_{s}^{*} = 0_{L}\) against the alternative hypothesis \(H_{1s}: \alpha_{s}^{*} \neq 0_{L}\) using a Bayesian Wald test. The test statistic is calculated as

\[
W_{s} = \tilde{a}_{s}^{*}\text{Var}(\alpha_{s}^{*})^{-1}\tilde{a}_{s}^{*},
\]

where \(\tilde{a}_{s}^{*}\) is the posterior mean of \(\alpha_{s}^{*}\), the posterior variance \(\text{Var}(\alpha_{s}^{*})\) is estimated by \(1/N\sum_{j=1}^{N}(\alpha_{1s}^{*} - \tilde{a}_{s}^{*})(\alpha_{1s}^{*} - \tilde{a}_{s}^{*})^{T}\), \(N\) is the total number of posterior samples, \(j\) is the index of \(j\)th posterior sample, and \(W_{s}\) is compared with a \(\chi^{2}\)-distribution with degree of freedom \(L\). If \(W_{s} > \chi^{2}_{L}(0.95)\), we conclude that the \(s\)th outcome does not have regression parameters in common with other outcomes.

4.2. Extended model B: one predictor has a distinct effect for each outcome

Suppose we have \(K^{*}\) outcomes remaining in our model. For a pre-specified predictor \(l\), we modify model (2.3) to allow the \(l\)th predictor \(x_{iL}\) to have a separate coefficient for each outcome \(k\). We modify \(\eta_{ik}\) for outcome \(k\) in model (2.3)

\[
\eta_{ik} = \gamma_{0k} + \gamma_{1x_{ik}}\frac{\alpha}{\|\alpha\|} + x_{ik}\phi_{k},
\]

where \(\phi_{k}, k = 1, \ldots, K^{*}\), is the departure from the common effect of the fixed effect of predictor \(l\) on outcome \(k\) and \(x_{ik}\) is the \(k\)th component of \(x_{i}\) for the \(k\)th subject; also recall that \(\alpha_{0}\) is the \(k\)th component of \(\alpha\). Define \(\phi = (\phi_{1}, \ldots, \phi_{K^{*}})\)' and \(\gamma_{1} = (\gamma_{11}, \ldots, \gamma_{1K^{*}})\)''. We assign \(\phi\) a normal prior with mean \(0_{K^{*}}\) and variance \(I_{K^{*}}\).

As with model B, we have an identifiability problem because \(\alpha_{0}\) and \(\phi_{k}\) can be replaced by \(\alpha_{0} = \alpha_{1}^{*}\) and \(\phi_{k} = \phi_{1}^{*} + (\gamma_{0}\alpha_{0}(1 - c))\|\alpha\|\) for an arbitrary constant \(c\), thus \(\gamma_{1}\) and \(\phi_{k}\) are confounded. We again use the Gram–Schmidt process to split \(\phi\) into two components, one component \(c^{*}\gamma_{1}\) is parallel to \(\gamma_{1}\), where \(c^{*} = \phi_{1}/\gamma_{1}\)' and the other component \(\phi_{k} = \phi_{k} - c^{*}\gamma_{1}\) are perpendicular to \(\gamma_{1}\).

We test the hypothesis \(H_{0k}: \phi_{k} = 0_{K^{*}}\) against the alternative hypothesis \(H_{1k}: \phi_{k} \neq 0_{K^{*}}\). The Wald test statistic is

\[
W_{\phi_{k}} = \phi_{k}^{*}\text{Var}(\phi_{k}^{*})^{-1}\phi_{k}^{*},
\]

where \(\phi_{k}^{*}\) is the posterior mean of \(\phi_{k}\), posterior variance \(\text{Var}(\phi_{k}^{*})\) is estimated by \(1/N\sum_{j=1}^{N}(\phi_{k,j}^{*} - \phi_{k}^{*})(\phi_{k,j}^{*} - \phi_{k}^{*})\) and \(W_{\phi_{k}}\) is compared with a \(\chi^{2}\)-distribution with degree of freedom \(K^{*}\). If \(W_{\phi_{k}} > \chi^{2}_{K^{*}}(0.95)\), we conclude that the \(k\)th predictor is not commonly predictive for all outcomes.

4.3. Final model

We fit extended model A for each outcome in turn and calculate each Wald test. We remove the outcomes with significant Wald test results from our COPE model. We then fit extended model B with the remaining outcomes for each predictor in turn and calculate each Wald test statistic as well. We remove predictors with significant Wald tests from the common set of predictors and these predictors are allowed to have separate coefficients for each outcome. Suppose, we have \(K^{*}\) outcomes left for the COPE model, \(L'\) predictors are common and \(L - L'\) predictors have outcome-specific regression coefficients. Without loss of generality, suppose the first \(L'\) predictors are common for all outcomes, denoted as \(x_{iL'} = (x_{i1}, \ldots, x_{iL'})\). Then our final model is

\[
\eta_{ik} = \gamma_{0k} + \gamma_{1s}x_{ik}\frac{\alpha}{\|\alpha\|} + \sum_{m=L'+1}^{L} x_{im}\alpha_{m} + \beta_{ik},
\]

where \(\alpha = (\alpha_{1}, \ldots, \alpha_{L})\) is the common regression coefficient vector, \(\alpha_{m}\) is the regression coefficient of predictor \(m\), \(m = L' + 1, \ldots, L\), for outcome \(k\), \(k = 1, \ldots, K^{*}\), \(\gamma_{0k}\), \(\gamma_{1s}\) and \(\beta_{ik}\) are the same as in model (2.3). If the Wald test for every outcome is significant, the simple solution is to not use the COPE model, though some B-type diagnostic model might still be appropriate. If all predictors are significant, we get the usual saturated model.

5. DATA ANALYSIS: TEENS LINKED TO CARE

We illustrate the methodology by examining baseline observations from the teens linked to care (TLC) study
on YPLH, which was conducted in eight adolescent clinical care sites in Los Angeles, New York, and San Francisco from 1995 to 1996 [37,38]. In this paper, we study outcomes that are subscales from the psychometric measure brief symptom inventory (BSI). The BSI consists of 53 items covering nine symptom dimensions: somatization, obsessive-compulsive disorder (OCD), interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism. Each item is given an integer-valued score ranging from 0 to 4.

We generate continuous, binary and Poisson variables from the 53 items by the following rules: continuous subscale scores are calculated by summing the values for the items included in that dimension and dividing by the number of items; binary variables are calculated from a single item score by setting the binary variable to 0 if the original item score is 0 else setting the binary variable to 1 for an item score of 1–4; and Poisson variables are calculated by summing the score of the items included in that dimension.

These are all commonly used distributions for items and subscales except for the Poisson distribution, which we think merits greater consideration than it has received in the past. For example, many items on the BSI are skewed with respondents choosing mostly 0, occasionally 1 and rarely 2 to 4. For an item scored on a 0 to 4 scale, a Poisson random variable with mean less than 1.27 has less than a 1 per cent chance of being 5 or larger, and thus the Poisson approximation to the sampling distribution has the potential to be surprisingly accurate; this must necessarily be confirmed for each item in the dataset. Summing several approximately Poisson-distributed random variables then produces a random variable that, in practice, is indistinguishable from a Poisson, even in the presence of modest correlation. Poisson regression is a commonly accepted model for non-negative integer-valued data; our model introduces the random effect to model both the correlation and the overdispersion, for when a non-negative integer outcome being modelled as Poisson has variance bigger than its mean.

We then apply the proposed COPE model on a set of generated outcomes that include continuous, binary and Poisson variables. Higher scores on every scale indicate a worse psychological state.

We use data collected at baseline interview in our analysis and start with a candidate model with eight outcomes, two treated as continuous subscales: depression and OCD; two treated as Poisson subscales: anxiety and somatization; and four dichotomized as binary items: BS11 (poor appetite), BS13 (temper uncontrollable), BS21 (people are unfriendly) and BS31 (avoid things that frighten you). No items are common between BS1 subscales and the four binary items belong to subscales not otherwise used in this analysis. The two continuous outcomes have long right tails, we transform each outcome as \( \log_2 (x + c) \), where \( c \) is the smallest non-zero value for that outcome, in our example, \( c = 1/6 \) for both outcomes. For response variables analysed on the log scale, the exponentiated regression coefficients can be interpreted as the multiplicative increase in the unlogged response per unit increase in the predictor.

The predictors are gender (73% male); age standardized to have mean 0 and standard deviation 1 by subtracting the mean 20.8 and then dividing by the standard deviation 2.1; food (87% yes), yes means that a subject can get enough food easily daily; finance (66% yes), yes means that a subject has the necessities to live comfortably and no means that the subject is struggling to survive and has difficulty paying bills; marijuana use in the past three months (79% yes); acquired immune deficiency syndrome (AIDS) symptoms (37% yes), yes means that the subject has physical health symptoms resulting from HIV infection; hard drug use (62% yes), yes means that the subject used at least one type of hard drug from stimulants, LSD, inhalants, coke, crack and heroin in the past three months; and attempted suicide (41% yes), yes means that the subject attempted suicide at least once prior to being enrolled in this study.

We complete a Bayesian prior specification of the model by following the recommendations from §3. We assign a normal prior with mean 0 and variance 9\( L_i \) to \( \alpha, \alpha \sim N(0,9L_i) \), here 9 can be changed to any arbitrary value. For the intercepts \( \gamma_0 \), for continuous outcomes, we use the centre of the range (max + min)/2 as the prior mean and 1/4 of the range (max − min)/4 as the prior standard deviation; for binary outcomes, we use 0 as the prior mean and 3.4 as the prior standard deviation; for Poisson data, we use 0 as the prior mean and 2.3 as the prior standard deviation. A guess of the ranges of log depression and OCD are both from −2.58 to 2.06. The centre of the range is −0.26 and the range divided by 4 is 1.16. Therefore, we set normal independent priors for \( \gamma_0 \) with means (−0.26, −0.26, 0, 0, 0, 0, 0, 0) and standard deviations (1.16, 1.16, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4).

We have one standardized continuous predictor and seven binary predictors, the range of \( x' \alpha / || \alpha || \) is then from −2 − \( \sqrt{7} \) to \( 2 + \sqrt{7} \) which is −4.6 to 4.6 and the length of the range is 9.2. A guess of the range of \( \eta_k \) is 5.2 for continuous outcomes after logarithm transformation, 13.8 for binary outcomes and 9.2 for Poisson outcomes. We calculate the prior standard deviation for scale parameters by dividing the range of \( \eta_k \) by 4 times the range of \( x' \alpha / || \alpha || \) and we get 0.14 for continuous outcomes, 0.38 for binary outcomes and 0.25 for count outcomes. We set normal independent priors for scale parameters \( \gamma_i \) with mean (0, 0, 0, 0, 0, 0, 0, 0) and standard deviations (0.28, 0.28, 0.76, 0.76, 0.76, 0.76, 0.50, 0.50).

We set a Wishart prior for the inverse of the covariance matrix for the random effects \( \beta \), with degrees of freedom 20 and we calculate the \( k \)th diagonal element by dividing \( 1 + K − 1 = 20 − 8 = 11 \) by a guess of the variance of \( \eta_k \), which are (5.2/4)\(^2\) = 1.7 for continuous outcomes, (13.8/4)\(^2\) = 11.9 for binary outcomes and (9.2/4)\(^2\) = 5.3 for count outcomes. We get the prior precision matrix \( \Lambda = diag(6.5,6.5,0,0,9,0,9,0,9,2,1.2) \).

We run the proposed candidate model in WinBUGS [39] and generate 500000 posterior samples, discarding the first 4000 samples as a burn-in. We examine autocorrelation plots and find that the autocorrelations for most parameters are close to 0 after 300 lags. The posterior density plots show that the posterior distributions for \( \alpha, \gamma_0 \) and \( \gamma_i \) are bell-shaped and unimodal.

We diagnose the candidate model by running extended model A, models (2.2\( \alpha \))–(4.1), for each
Table 1. Model diagnostics for checking whether an outcome or a predictor belongs in the COPE model. Left three columns: the Bayesian Wald statistic $W_a$ in extended model A is used to test whether an outcome has a separate set of regression coefficients different from other outcomes. The $p$-value next to $W_a$ comes from comparing $W_a$ with $\chi^2$, a $\chi^2$-distribution with 8 d.f. Anxiety is the only outcome whose Wald statistic is significant. We exclude anxiety from the COPE model. Right three columns: Wald statistics $W_\phi$ are from extended model B and are used to test whether a predictor has a different regression coefficient for each outcome. The $p$-value next to $W_\phi$ comes from comparing $W_\phi$ with a $\chi^2$ distribution. Suicide attempt is the only predictor whose Wald statistic is significant.

<table>
<thead>
<tr>
<th>outcome</th>
<th>$W_a$</th>
<th>$p$-value</th>
<th>predictor</th>
<th>$W_\phi$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>depression</td>
<td>10.2</td>
<td>0.25</td>
<td>male</td>
<td>8.5</td>
<td>0.29</td>
</tr>
<tr>
<td>obsessive compulsive</td>
<td>9.8</td>
<td>0.28</td>
<td>age</td>
<td>4.2</td>
<td>0.76</td>
</tr>
<tr>
<td>BS11 (poor appetite)</td>
<td>5.5</td>
<td>0.70</td>
<td>food</td>
<td>8.1</td>
<td>0.33</td>
</tr>
<tr>
<td>BS13 (temper uncontrol)</td>
<td>10.3</td>
<td>0.24</td>
<td>finance</td>
<td>4.0</td>
<td>0.78</td>
</tr>
<tr>
<td>BS21 (people not friendly)</td>
<td>5.4</td>
<td>0.72</td>
<td>marijuana</td>
<td>6.2</td>
<td>0.52</td>
</tr>
<tr>
<td>BS31 (avoid frighten)</td>
<td>9.0</td>
<td>0.34</td>
<td>hard drugs</td>
<td>9.6</td>
<td>0.21</td>
</tr>
<tr>
<td>anxiety</td>
<td>15.6</td>
<td>0.05</td>
<td>AIDS symptoms</td>
<td>6.8</td>
<td>0.45</td>
</tr>
<tr>
<td>somatization</td>
<td>13.7</td>
<td>0.09</td>
<td>suicide attempt</td>
<td>14.7</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 2. Posterior summaries for common predictors $\alpha$, scale parameters $\gamma_1$ and intercept $\gamma_0$ in final model. Age is standardized with mean 0 and standard deviation 1, other covariates are 0–1 variables; mean and standard deviation are the posterior sample mean and standard deviation for each parameter, respectively. Column headings 2.50% and 97.50% are the 2.5 percentile and 97.5 percentile, respectively, and $\% \geq 0 = P(\text{parameter} \geq 0 \mid Y)$ is an estimate of the posterior probability that the parameter is positive.

<table>
<thead>
<tr>
<th>predictor</th>
<th>parm</th>
<th>mean</th>
<th>s.d.</th>
<th>2.50%</th>
<th>97.50%</th>
<th>$% \geq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>scale parameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>$\gamma_1$</td>
<td>0.58</td>
<td>0.16</td>
<td>0.27</td>
<td>0.89</td>
<td>1.00</td>
</tr>
<tr>
<td>obsessive compulsive</td>
<td>$\gamma_2$</td>
<td>0.71</td>
<td>0.16</td>
<td>0.39</td>
<td>1.03</td>
<td>1.00</td>
</tr>
<tr>
<td>BS11 (poor appetite)</td>
<td>$\gamma_3$</td>
<td>1.42</td>
<td>0.37</td>
<td>0.71</td>
<td>2.18</td>
<td>1.00</td>
</tr>
<tr>
<td>BS13 (temper uncontrol)</td>
<td>$\gamma_4$</td>
<td>0.47</td>
<td>0.36</td>
<td>−0.24</td>
<td>1.17</td>
<td>0.91</td>
</tr>
<tr>
<td>BS21 (people not friendly)</td>
<td>$\gamma_5$</td>
<td>0.87</td>
<td>0.34</td>
<td>0.22</td>
<td>1.56</td>
<td>1.00</td>
</tr>
<tr>
<td>BS31 (avoid frighten)</td>
<td>$\gamma_6$</td>
<td>0.71</td>
<td>0.38</td>
<td>−0.05</td>
<td>1.45</td>
<td>0.97</td>
</tr>
<tr>
<td>somatization</td>
<td>$\gamma_7$</td>
<td>0.71</td>
<td>0.15</td>
<td>0.42</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td>intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>$\gamma_0$</td>
<td>−0.35</td>
<td>0.17</td>
<td>−0.67</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>obsessive compulsive</td>
<td>$\gamma_2$</td>
<td>−0.07</td>
<td>0.20</td>
<td>−0.44</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td>BS11 (poor appetite)</td>
<td>$\gamma_3$</td>
<td>−0.10</td>
<td>0.42</td>
<td>−0.88</td>
<td>0.76</td>
<td>0.39</td>
</tr>
<tr>
<td>BS13 (temper uncontrol)</td>
<td>$\gamma_4$</td>
<td>−0.27</td>
<td>0.27</td>
<td>−0.69</td>
<td>0.37</td>
<td>0.19</td>
</tr>
<tr>
<td>BS21 (people not friendly)</td>
<td>$\gamma_5$</td>
<td>−0.13</td>
<td>0.31</td>
<td>−0.43</td>
<td>0.79</td>
<td>0.64</td>
</tr>
<tr>
<td>BS31 (avoid frighten)</td>
<td>$\gamma_6$</td>
<td>−0.80</td>
<td>0.31</td>
<td>−1.38</td>
<td>−0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>somatization</td>
<td>$\gamma_7$</td>
<td>1.11</td>
<td>0.20</td>
<td>0.73</td>
<td>1.51</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Outcome in turn to test whether the outcome has a separate set of regression coefficients. This results in eight extended models corresponding to the eight outcomes. We calculate the Bayesian Wald test statistic for each model and compare it with $\chi^2(0.95) = 15.51$, where the degree of freedom is equal to the number of predictors in the model. Results show that the Wald test statistic is significant for anxiety ($p$-value = 0.048), but not for any of the other outcomes. We exclude the outcome anxiety then run extended model B, models (2.2a)–(2.2c) and (4.3) for each predictor in turn to test whether the predictor is not commonly predictive for all outcomes, but has a different regression coefficient for each outcome, resulting in another eight extended models corresponding to the eight predictors. We calculate the Bayesian Wald test statistic for each model and compare it with $\chi^2(0.95) = 14.07$, where the degree of freedom is the number of outcomes in the model, which is 7 now because we excluded anxiety. Results show that the Wald test statistic is significant for suicide attempt ($p$-value = 0.040) but not for the other predictors. The Bayesian Wald test statistics and $p$-values for these two sets of extended models are shown in table 1.
Table 3. Regression coefficients. There are three columns for each outcome. Columns give the posterior mean, posterior standard deviation and percentage of positive posterior samples. We fit anxiety separately and use a relaxed COPE model for the other 7 outcomes. In the COPE model, For the first 7 common predictors, we calculate $a_l k g$ from the posterior samples of $a_l s$ and $g 1 k$ then draw inference from the calculated posterior samples of $a_l g 1 k$ and we have the outcome-specific regression coefficient for suicide attempt, the one predictor that has a separate regression coefficient for each outcome.

<table>
<thead>
<tr>
<th>pred</th>
<th>depression</th>
<th>OCD</th>
<th>BS11</th>
<th>BS13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>% ≥ 0</td>
<td>mean</td>
</tr>
<tr>
<td>male</td>
<td>0.03</td>
<td>0.12</td>
<td>0.59</td>
<td>0.03</td>
</tr>
<tr>
<td>age</td>
<td>0</td>
<td>0.04</td>
<td>0.54</td>
<td>0.01</td>
</tr>
<tr>
<td>food</td>
<td>-0.34</td>
<td>0.14</td>
<td>0</td>
<td>-0.42</td>
</tr>
<tr>
<td>financial</td>
<td>-0.19</td>
<td>0.10</td>
<td>0.02</td>
<td>-0.24</td>
</tr>
<tr>
<td>marijuana</td>
<td>0.10</td>
<td>0.11</td>
<td>0.83</td>
<td>0.12</td>
</tr>
<tr>
<td>hard drugs</td>
<td>0.03</td>
<td>0.12</td>
<td>0.58</td>
<td>0.04</td>
</tr>
<tr>
<td>AIDS symptoms</td>
<td>0.33</td>
<td>0.11</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>suicide attempt</td>
<td>0.61</td>
<td>0.13</td>
<td>1.00</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pred</th>
<th>som</th>
<th>anx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d. % ≥ 0</td>
</tr>
<tr>
<td>male</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>age</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>food</td>
<td>-0.51</td>
<td>0.25</td>
</tr>
<tr>
<td>financial</td>
<td>-0.30</td>
<td>0.19</td>
</tr>
<tr>
<td>marijuana</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>hard drugs</td>
<td>0.03</td>
<td>0.19</td>
</tr>
<tr>
<td>AIDS symptoms</td>
<td>0.50</td>
<td>0.22</td>
</tr>
<tr>
<td>suicide attempt</td>
<td>0.20</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Our final model has anxiety by itself in a separate model and a relaxed COPE model for the other seven outcomes, and eight predictors among which suicide attempt has an outcome-specific effect and the other seven predictors constitute a single linear predictor commonly predictive for the seven remaining outcomes. Posterior summaries for the seven common regression coefficients $\alpha$ and eight scale parameters $\gamma_i$ are shown in Table 2. Having enough food and a good financial situation imparts negative effects on all seven outcomes. Using marijuana and having AIDS symptoms have positive effects on all seven outcomes. The magnitude of the effects is modified by the scale parameter for each outcome. The scale parameter for the depression and OCD subscales is 0.58 and 0.71, respectively, which means that 1 unit change in the single linear predictor would associate with a change of $2^{0.58} = 1.49$ units in depression and a change of $2^{0.71} = 1.64$ for the OCD subscale because we used a base 2 logarithm transformation for these two continuous variables prior to analysis. The scale parameters for binary outcomes B11, B13, B21 and B31 are 1.42, 0.47, 0.87 and 0.71, respectively, which means that 1 unit change in the single linear predictor would associated with $e^{1.42} = 4.13$, $e^{0.47} = 1.59$, $e^{0.87} = 2.39$ and $e^{0.71} = 2.04$ odds ratio change for B11, B13, B21 and B31. The scale parameter for Poisson outcome somatization is 0.71, indicating a one unit change in linear predictor is associated with $e^{0.71} = 2.04$ unit change in somatization.

Table 3 shows posterior summaries for the regression coefficients of each predictor for each outcome. We fit a separate model for anxiety and a relaxed COPE model for the other seven outcomes. For the first seven common predictors, we calculate $\alpha_i \gamma_{ik}$ from the posterior samples of $\alpha_i$ and $\gamma_{ik}$ then draw inference from the calculated posterior samples of $\alpha_i \gamma_{ik}$. We have the outcome-specific regression coefficient for suicide attempt, the one predictor that has a separate regression coefficient for each outcome. Suicide attempt has significant positive effects on depression, OCD, B31 and somatization. Food and finance have significant negative effects on anxiety while hard drugs, AIDS symptoms and suicide attempt have significant positive effects. We compare the standard deviation of the regression coefficients in our final relaxed COPE model with that in the saturated model and find that on average, the standard deviation in our final model is 29 per cent less than that in the saturated model. This is evidence that compared with the saturated model, the COPE model has greater efficiency.

6. DISCUSSION

We began with a candidate COPE model that assumed all outcomes were similar and shared one common linear predictor. We then fit flexible diagnostic models to identify outcomes that did not belong to the COPE model, and which were then excluded from the final model. The model with one linear predictor is a one cluster model. When we have a large number of outcomes, more than one linear predictor may be needed to predict all outcomes. We may expand our one cluster model to a multi-cluster model by clustering outcomes and introducing cluster-specific regression coefficients for each cluster [40].

We can use a graphical method to identify which outcomes belong to the same cluster and how many clusters we shall have. We first run a univariate generalized linear model for each outcome $y_k$ and get estimates $\hat{\alpha}_k = (\hat{\alpha}_{k1}, \ldots, \hat{\alpha}_{kL})$ of regression coefficients omitting the intercept. We normalize $\hat{\alpha}_k$ by dividing by its Euclidean norm, $\hat{\alpha}_0 = \hat{\alpha}_k / \sqrt{\hat{\alpha}_k^t \hat{\alpha}_k}$, so that $\hat{\alpha}_0$ has length 1 and can be compared across outcomes. We plot a profile plot of the $\hat{\alpha}_k$, plotting points $(j, \hat{\alpha}_k)$ and drawing line segments connecting point $(j, \hat{\alpha}_k)$ to $(j+1, \hat{\alpha}_{k+1})$, and inspect the plot to identify potential outcome clusters.

Finally, we can adapt the COPE model to longitudinal mixed outcome data by introducing two latent variables, one to model the correlation among outcomes at the same time and the other to model the correlation among outcomes over time. Consideration of the COPE model when it cannot be written as a linear model, as in model (2.3), could be an area of future work.

In this paper, we proposed a Bayesian Wald test to determine whether one outcome or one predictor belongs to the COPE model. There are alternative methods for model selection. The Bayes factor is a widely used statistic in which prior and posterior information are combined in a ratio that provides evidence in favour of one model versus another. We can calculate a Bayes factor as the marginal likelihood of the candidate model divided by the marginal likelihood of a diagnostic type A model or type B model, where the marginal likelihood of a model is the probability of the data with all the model parameters integrated out. We can also use Akaike information criterion (AIC), Bayesian information criterion (BIC) or deviance information criterion (DIC) to do model selection. In a Bayesian framework, DIC is easily calculated as a linear model, as in model (2.3), could be an area of future work.

Weiss was supported in part by the Center for HIV Identification, Prevention and Treatment Services, NIH/NIMH P30MH58107. Suchard was supported in part by a John Simon Guggenheim Memorial Foundation Fellowship and a research gift from Google.

REFERENCES

Mixed outcomes with common effects  R. E. Weiss et al