Published in final edited form as:

Stat Med. 2009 June 15; 28(13): 1793–1804. doi:10.1002/sim.3589.

Common predictor effects for multivariate longitudinal data

Juan Jia and Robert E. Weiss*

Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095-1772 U.S.A.

SUMMARY

Multivariate outcomes measured longitudinally over time are common in medicine, public health, psychology and sociology. The typical (saturated) longitudinal multivariate regression model has a separate set of regression coefficients for each outcome. However, multivariate outcomes are often quite similar and many outcomes can be expected to respond similarly to changes in covariate values. Given a set of outcomes likely to share common covariate effects, we propose the Clustered Outcome COMmon Predictor Effect (COPE) model and offer a two step iterative algorithm to fit the model using available software for univariate longitudinal data. Outcomes that share predictor affects need not be chosen a priori; we propose model selection tools to let the data select outcome clusters. We apply the proposed methods to psychometric data from adolescent children of HIV+ parents.

Keywords

Clustering; Common Effect; Dimension Reduction; Hierarchical Linear Model; HIV; Multivariate Regression; Model Selection

1. Introduction

Longitudinal studies in medicine, public health, psychology and sociology measure multiple outcomes repeatedly over time. Multivariate longitudinal outcomes are often highly correlated and may be affected similarly by covariates. We might find that as the covariates X_1 , X_2 and X_3 increase, so do outcomes Y_1 , Y_2 and Y_3 . Rather than fit separate models with separate linear predictors for all three outcomes, we propose that outcomes with similar relationships to covariates be combined into a single model with a single linear predictor that predicts all outcomes.

Numerous multivariate longitudinal data models have been proposed in recent years [1, 2, 3, 4, 5, 6, 7, 8, 9]. For univariate longitudinal data, main interest lies in covariate effects. For multivariate outcomes, interest often lies in the covariances that measure the interrelation among outcomes [10, 11, 12, 13] and testing global effects and estimating common effects on multivariate outcomes [14, 15, 16]. O'Brien [17] proposed a global test for a common dose effect for multiple outcomes. Sammel, Ryan, and Legler [18] proposed a latent variable model where the multiple outcomes were summarized by a single latent variable and a univariate linear regression model was used to assess common covariate effects. Lin and coauthors [19] considered a scaled linear mixed model for multiple outcomes that standardized each outcome by its standard deviation to estimate the common effect of a single covariate of interest. In our paper, we are interested in a middle area where we

analyze multiple outcomes, yet primary interest lies in the relationship of outcomes to predictors.

A linear predictor is a linear combination of covariates. We define two or more outcomes as *similar* if they relate to covariates similarly which we define to mean that a single linear predictor is all that is needed to predict the similar outcomes. *Dissimilar* outcomes need different linear predictors. Similar outcomes may be identified *a priori* from expert knowledge, or preferably in many circumstances, we can let the data determine which outcomes are similar. Similar outcomes belong to a single cluster of outcomes in our model.

We define the Clustered Outcome COMmon Predictor Effect (COPE; CO stands for both Clustered Outcome and COMmon) model for multivariate outcomes as a model where similar outcomes share a single linear predictor while dissimilar outcomes have different linear predictors. Outcome-specific link functions connect the several similar outcomes to the one linear predictor.

The *saturated model* is the usual multivariate model where each outcome has its own outcome-specific coefficients. The COPE model will have greater efficiency in parameter estimation than the saturated model. Suppose there are 5 outcomes and 10 predictors, then the saturated model estimates $5 * 10 = 50$ fixed effect parameters and 5 intercepts for a total of 55 parameters. If all outcomes are similar, we need to estimate 19 total parameters, 10 fixed effects, 5 intercepts and 4 scale parameters. To avoid identifiability problems, we force one of the scale parameters equal to one in each cluster. If outcomes are grouped into two clusters, we would need to estimate 28 parameters, $2 * 10 = 20$ fixed effects parameters, 5 intercepts and 3 scale parameters. In general, with K outcomes, G clusters and L covariates not including the intercept, there are K intercepts, $K - G$ scale parameters and GL coefficients for a total of $2K + G(L - 1)$ parameters, which can be much less than the $K(L + 1)$ coefficients for the saturated model, particularly when G is small.

The COPE model increases scientific understanding of the outcomes by identifying sets of outcomes that change similarly as covariates change and it eases the difficulty of interpreting large numbers of parameters. If a covariate x_l is a significant predictor for the g^{th} cluster then that covariate is a significant predictor for all outcomes in the cluster and this takes only a single hypothesis test.

This paper is motivated by a study of parents living with Human Immunodeficiency Virus (HIV) and their adolescent children. HIV+ parents were recruited from New York City's Division of AIDS Services along with their adolescent children and were followed for up to 7 years. Follow-up interviews were conducted every three months in the first two years and every six months thereafter. Age, female (coded yes=1/no=0), time of visit, parental drug status, parental alcohol use, season, brief symptom inventory (BSI) sub-scales, of which we analyze Anxiety and Somatization, and several coping style variables, of which we analyze Positive Action and Social Support, were recorded at each interview along with many other variables. We find the best way to cluster the outcomes and estimate common predictor (female, age, parental drug status, season and so on) effects for each outcome cluster.

This paper is organized as following: Section 2 develops the Clustered Outcome COMmon Predictor Effects (COPE) model in the situation that all outcomes belong to one cluster, then more generally for G clusters. Section 3 explains the two step iterative algorithm to fit the model and model selection tools for specifying the outcome clusters. Section 4 applies the method to the data of adolescent children with HIV+ parents. The paper closes with a discussion in section 5.

2. The Clustered Outcome Common Predictor Effects (COPE) model

We observe K continuous outcomes and L covariates on subject i , $i = 1, \dots, n$, at times t_{ij} , $j = 1, \dots, J_i$. Let y_{ijk} denote the k^{th} outcome for subject i at time t_{ij} , $k = 1, \dots, K$, and define $y_{ij} = (y_{ij1}, \dots, y_{ijK})'$; x_{ijl} is the value of the l^{th} predictor for subject i at time j , $l = 1, \dots, L$; $x_{ij} = (x_{ij1}, \dots, x_{ijL})'$ and x_{ij} does not include an intercept. Covariates are the same for all outcomes. The saturated multiple outcomes longitudinal model is

$$y_{ijk} = \gamma_{0k} + x_{ij}'\alpha_k + \varepsilon_{ijk} \quad (1)$$

where γ_{0k} is the intercept for outcome k , α_k is an $L \times 1$ vector of fixed effect coefficients for the k^{th} outcome, ε_{ijk} are random errors, $\varepsilon_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijK})'$, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ_i})'$, $\varepsilon_i \sim N(0, \Sigma(\theta))$, $\Sigma(\theta)$ models covariances among the multiple outcomes and repeated measures, and θ is a vector of unknown parameters for the components of the covariance matrix.

First we specify a single cluster COPE model, meaning that a single linear predictor $x_{ij}\alpha$ predicts all outcomes. However, outcomes are measured on different scales and it is necessary to allow each outcome to have its own link function to connect with the linear predictor. The model is

$$E(y_{ijk}) = u_{ijk} \quad u_{ijk} = \psi_k(x_{ij}\alpha) = \gamma_{0k} + \gamma_{1k}x_{ij}\alpha \quad (2)$$

where u_{ijk} is the expectation of the k^{th} outcome of subject i at time t_{ij} and the link function $\psi_k(\cdot)$, here assumed linear, links the mean to the linear predictor using intercept γ_{0k} and scale parameter γ_{1k} . In (2), $\alpha = (\alpha_1, \dots, \alpha_L)'$ is the $L \times 1$ vector of fixed common predictor effects.

In practice, each outcome will belong to one of G clusters with outcomes in each cluster sharing a common predictor effect. Here we assume the clustering is known; model selection tools to identify similarity among outcomes are discussed in the next section. There is an additional nested subscript compared to (2); we introduce subscript g to index clusters, with k now indexing outcomes within cluster. In cluster g , $g = 1, \dots, G$, there are k_g outcomes, $\sum_{g=1}^G k_g = K$. The k^{th} outcome in cluster g for subject i at time t_{ij} is y_{ijgk} and we model

$$E(y_{ijgk}) = u_{ijgk} = \psi_{gk}(x_{ij}\alpha_g) = \gamma_{0gk} + \gamma_{1gk}x_{ij}\alpha_g \quad (3)$$

where γ_{0gk} is the intercept for the k^{th} outcome, $k = 1, \dots, k_g$ in the g^{th} cluster, $g = 1, \dots, G$, γ_{1gk} is the scale parameter for the k^{th} outcome in the g^{th} cluster, α_g is the $L \times 1$ vector of fixed common effects for all outcomes in the g^{th} cluster, $\alpha_g = (\alpha_{g1}, \dots, \alpha_{gL})'$ and α_{gl} is the regression coefficient for the l^{th} predictor in the g^{th} cluster.

3. Model fitting and selection

We separate estimation into two parts, (i) estimation given a set of clusters and (ii) identification of a best model, that is, identification of the best way to cluster outcomes.

3.1. Estimation given a set of clusters

Given a clustering of outcomes, we estimate parameters (γ, α, θ) , where link function parameters are $\gamma = (\gamma_0', \gamma_1')'$ (K intercepts γ_0 and K scaling parameters γ_1) where $\gamma_s = (\gamma_{s11}, \dots, \gamma_{s1k_1}, \dots, \gamma_{sG1}, \dots, \gamma_{sGk_G})'$, $s = 0, 1$, and fixed effects parameters are $\alpha = (\alpha_1', \dots, \alpha_G')'$, and regression coefficients in the g^{th} cluster are $\alpha_g = (\alpha_{g1}, \dots, \alpha_{gL})'$, $g = 1, \dots, G$.

We develop a two step iterative algorithm for fitting (3) using already existing algorithms for maximum likelihood fitting of multivariate longitudinal data given a particular clustering of the outcomes. We fit (3) by first conditioning on $\gamma = (\gamma'_0, \gamma'_1)'$ and maximizing with respect to (α, θ) . Next we condition on α and maximize the likelihood with respect to (γ, θ) . We iterate till convergence by repeatedly calling existing univariate longitudinal statistical software, in our case the MIXED procedure in SAS (SAS Institute, Cary, NC).

To start the algorithm, we first get an initial value $\alpha^{(0)}$ for α . We estimate $\alpha^{(0)}$ by fitting the usual multivariate longitudinal linear model $y_{ijgk} = \gamma_{0g} + x_{ij}\alpha_g + \varepsilon_{ijgk}$ with G clusters but with a simplified link function where the intercepts $\gamma_{0gk} \equiv \gamma_{0g}$ are all equal and all $\gamma_{1gk} \equiv 1$.

Suppose we have executed $t - 1$ previous iterations, now we are starting the t^{th} iteration.

The steps are

- A.** Update link. Define $\hat{V}_{ijg}^{(t-1)} = x_{ij}\hat{\alpha}_g^{(t-1)}$. Estimate $\hat{\gamma}_{0gk}^{(t)}$ and $\hat{\gamma}_{1gk}^{(t)}$ by fitting the multivariate longitudinal linear model

$$y_{ijgk} = \gamma_{0gk} + \gamma_{1gk}\hat{V}_{ijg}^{(t-1)} + \varepsilon_{ijgk}, \quad (4)$$

fixing $\hat{\alpha}_g^{(t-1)}$ and $\hat{V}_{ijg}^{(t-1)}$. In each cluster, to avoid a lack of identifiability, we force one scale parameter equal to 1. Without loss of generality, we set the scale parameter for the last outcome in each cluster to be 1, where last is defined alphabetically or with other arbitrary ordering.

- B.** Update the common α_g . Define $\hat{W}_{igk}^{(t)} = \hat{\gamma}_{1gk}^{(t)}x_i$ and $z_{ijgk}^{(t)} = y_{ijgk} - \hat{\gamma}_{0gk}^{(t)}$. Estimate $\alpha^{(t)}$ s by fitting the multivariate longitudinal model

$$z_{ijgk} = \hat{W}_{1gk}^{(t)}\alpha_g + \varepsilon_{ijgk} \quad (5)$$

with no intercept.

The maximized likelihood increases after each iteration and after each step within an iteration. In calling SAS Proc Mixed, we update the variance parameters θ at each step of the algorithm. Many methods for checking convergence are possible. We record the maximized likelihood and repeat steps (A) and (B) until the difference of maximized likelihoods in two consecutive iterations is less than a pre-specified small number.

We calculate standard errors of the estimated α and γ as the square root of the diagonal elements of the negative inverse of the observed information matrix. We don't use the output from either step in our algorithm. Formulas are given in the Appendix.

3.2. Choosing a best set of clusters

A best partition of outcomes into disjoint clusters can be determined by the data. We look for a clustering of similar outcomes that give the largest likelihood and that is not significantly worse than the saturated model. For K outcomes, there are $b(K)$ ways to partition the K outcomes, where $b(K)$ is the Bell number [20]. There are $b(4) = 15$ candidate models for four outcomes. We use commas to separate clusters within a model and semicolons to separate models, and use the numbers 1–4 to represent the four outcomes: the 15 models are 1,2,3,4; 12,3,4; 13,2,4; 14,2,3; 1,23,4; 1,24,3; 1,2,34; 123,4; 124,3; 12,34; 134,2; 13,24; 14,23; 1,234; 1234. Model 1,2,3,4 is the saturated model where each outcome

is its own cluster. Model 1234 is the single cluster COPE model where all outcomes are assumed similar to each other. The other 13 models have 2 or 3 clusters.

The number of models $b(K) = \sum_{r=1}^K a(K, r)$, where $a(K, r)$, the Stirling number of the second kind, is the number of ways to partition K outcomes into r clusters. Based on the recursion $a(n+1, r) = r \times a(n, r) + a(n, r-1)$ [21], knowing all partitions $a(n, r)$ and $a(n, r-1)$, we can generate all partitions of $n+1$ outcomes into r clusters using induction. Suppose we have partitioned n outcomes into r clusters, the $(n+1)^{st}$ outcome goes in to one of the existing $a(n, r)$ clusters, or we have partitioned n outcomes into $r-1$ clusters and the $(n+1)^{st}$ outcome becomes a new cluster, the $(r)^{th}$ cluster combined with the $(r-1)$ clusters in each partition of size $a(n, r-1)$. We illustrated model choice using AIC backed up by formal hypothesis testing against the saturated model.

4. Application

We analyze data on the psychometric outcomes of adolescent children of HIV+ parents introduced in section 1. Adolescents' data were collected every three months in the first two years and every six months thereafter for a total of up to 7 years. We analyze four outcomes, two are coping style subscales: 1=Social Support and 2=Positive Action, while 3=Anxiety and 4=Somatization, are BSI subscales. We number the variables from one to four to enable shorthand for identifying the different clusterings. The four outcome variables have long right tails, and we transform the outcomes as $\log_2(x+c)$ where c is the smallest non-zero outcome.

The predictors are female; age (years); season trichotomized as spring (Mar–Jun), summer (Jul–Oct) and winter (Nov–Feb); parental hard drug use trichotomized as non-user, using-user and non-using user; and time. Time is modeled as a linear spline with knots at 18 and 36 months; the three variables are titled Month, Month18 and Month36. A parent is a hard drug non-user if the parent never reported hard drug use during the study; a hard drug user is a parent who reported hard drug use at some visit during the study. Adolescents had each visit classified as (i) a non-user visit if their parent was a non-user; (ii) a using-user visit if the parent was a hard drug user and the parent reported hard drug use in the three months previous to the observation; or (iii) a non-using user visit if the parent was a hard drug user who did not report hard drug use in the previous three months.

Correlations among the four outcomes are given in Table I. The two BSI variables 3=Anxiety and 4=Somatization are highly correlated ($\rho = .75$) and the two coping style variables 1=Social Support and 2=Positive Action are moderately correlated ($\rho = .49$). The two coping style variables have low cross correlations with the two BSI variables (all $\rho < .25$).

We model the residuals ε_i with a multivariate random intercept model [1, chapter 13], one random intercept per outcome per subject. We write $\varepsilon_{ijk} = \beta_{ik} + \delta_{ijk}$, with $\beta_i \sim N_K(0, D)$ a K -vector with elements β_{ik} and $\delta_{ij} \sim N_K(0, V)$ where $\delta_{ij} = (\delta_{ij1}, \dots, \delta_{ijK})'$ and both D and V are unstructured covariance matrices.

We used the two step iterative algorithm to fit each model and compare the likelihood to the saturated model. We seek models that have the fewest clusters and at the same time are not significantly different from the saturated model. Results are summarized in Table II, the column difference is the likelihood ratio test statistic between each model and the saturated model, model 1,2,3,4; the degrees of freedom and p -values using the nominal χ^2 comparison distribution are listed in neighboring columns. The four columns labeled gamma give the estimated scale parameters γ_{1gk} for each model; the γ s are in the same order as the outcomes

are listed in the Model column. For example, in model 134,2, γ_1 is the scale parameter for outcome 1, γ_2 is the scale parameter for outcome 3, γ_3 is the scale parameter for outcome 4 and γ_4 is the scale parameter for outcome 2.

Model 134,2 with two clusters is best according to AIC. It is not significantly worse than the saturated model (p -value = .17). Models 13,2,4; 14,2,3; and 1,2,34 are also good models that are also not significantly worse than the saturated model (p -values = .22, .07, .46), but they have more parameters and one more cluster than model 134,2. Model 134,2 is nested in all three and all are not significantly better than model 134,2, thus we settle on model 134,2.

We also report BIC, though it is known to pick models with too few parameters as compared with AIC. BIC picks the model 1234, but this model is significantly worse than the model 1,2,3,4 and we discard it from further consideration.

The model selection results suggest that 1=Social Support, 3=Anxiety and 4=Somatization share a common set of predictor effects but 2=Positive Action has its own set of predictor effects. Based on the correlations in Table I, we might have expected the model 12,34 to have been best; actually model 12,34 is significantly worse than the saturated model. In fact, every model that clusters outcome 2 with any other outcome is significantly worse than the saturated model by the likelihood ratio test. This illustrates that outcome correlation and outcome similarity are different concepts and can lead to different clustering of outcomes even in the same data set.

The scale parameters γ_{1gk} for 1=Social Support and 3=Anxiety are .38 (se = .087) and .88 (se = .099) with 4=Somatization and 2=Positive Action having $\gamma \equiv 1$. Predictor changes associated with a one unit change in 4=Somatization are associated with a .38 unit change in 1=Social Support and a .88 unit change in 3=Anxiety. Both estimated γ_{1gk} s are significantly different from zero, indicating that both outcomes are indeed related to the linear predictor. The estimate of γ_{1gk} is not significantly different from 1, suggesting that it is possible that 3=Anxiety and 4=Somatization are affected equally by changes in covariates.

Table III shows parameter estimates, standard errors and p -values for $\gamma_{1gk}\alpha_g$ from the saturated model 1,2,3,4 and from our preferred model 134,2. We separate estimates, standard errors and p -values into three sub-sections of the table. Because we calculate $\gamma\alpha$ for each outcome, the p -values are slightly different for each outcome. If we depend on the results for α_g alone to report significance, which we would in a paper, those results are given in the columns for outcomes 2=Positive Action and 4=Somatization since their γ_{1gk} are constrained to equal 1.

In the saturated model, coefficient estimates for 3=Anxiety and 4=Somatization are close for statistically significant predictors including month, female, summer, month18 and these estimates are close to the estimated α for cluster 134. This is consistent with the scale parameter estimate $\gamma = .88$ for 3=Anxiety and $\gamma \equiv 1$ for 4=Somatization. 1=Social Support has a $\gamma_1 = .38$, and its estimates are therefore smaller and its standard errors (SE) are smaller compared to those for 3=Anxiety.

For outcome 2=Positive Action, age is important, however age is not significantly associated with outcomes 1, 3 or 4. Conversely, female gender is associated with the clustered outcomes 1,3 and 4 but is not significantly associated with outcome 2. Thus the differences in the clusters appears to be driven heavily by their relationships to covariates age and female gender.

The standard errors (SEs) for the COPE model are smaller than the corresponding SEs from the saturated model for all outcomes and coefficients except age. The average ratio of

standard errors COPE model to saturated model is .90 for a reduction in standard error of 10%; ignoring age, the average ratio of SEs is .76 for a reduction of 24%. This will be true in general as information from multiple outcomes estimates the coefficients for a single cluster of similar outcomes.

5. Discussion

We have described methodology for parameter estimation and model selection for the COPE model using maximum likelihood. Our model can be applied to both longitudinal and cross sectional data. Our original inspiration for this model is the longitudinal data set described in this example. A useful feature of the methodology is that inferences do not depend on the parameterization of the predictors; linear transformations of the predictors lead to the same choices and conclusions. While we use maximum likelihood for our estimation procedure, users who preferred to could use generalized estimating equations [22] or quasi-least squares [23] or other procedure. Like many statistical procedures, our approach is exploratory; a confirmatory approach might use one data set to select a clustering and a second data set could be used to specifically test the outcome clustering found by the first model. The second data set could result from an independent sample or it could be created by splitting an original larger sample into an estimation data set and a formal test data set.

While we did not emphasize the point, our model and likelihood based approach accommodates missing data that is missing at random [24]. Single outcomes at a time may be missing or all outcomes at a given time may be missing. Missing data is appropriately accounted for in our formulas for estimating the information matrix given in the appendix.

Lin et al [19] propose a somewhat related random effects model for cross sectional data. Our models differ in several major features. We use the data to choose our clusters, they assume outcomes under analysis are all in the single cluster. In their model, a single covariate has an effect that is constant across outcomes up to a scaling parameter, all other covariates are assumed to have different effects on different outcomes; within a cluster, we assume all covariates have similar impacts on the outcome up to the scaling parameter. Lin et al treat cross sectional data while we treat longitudinal data – this has a major impact on our approaches to scaling the impacts of the covariates, for example we conceptualize outcome specific link functions to connect the linear predictor to each outcome in a cluster. In general models for longitudinal data, there is not a single marginal variance for any outcome; the marginal variance is a function of time in any non-trivial model with a random time effect. We place our scaling parameters solely inside the linear predictor and we estimate them as part of the linear predictor; they are not related to the covariance parameters. This simplifies the algebra in the sense that unlike in Lin et al (2000) the covariance parameters are asymptotically independent of the linear predictor parameters for normal error distributions.

Acknowledgments

The authors thank Mary Jane Rotheram for permission to use the Project TALC data.

Contract/grant sponsor: Weiss was supported by the Center for HIV Identification, Prevention, and Treatment Services, NIH/NIMH; contract/grant number: P30MH58107

APPENDIX

Estimating the covariance matrix for $\hat{\alpha}$ and $\hat{\beta}$

We calculate the observed information matrix to estimate the covariance matrix for all fixed effect parameters in a K outcome G cluster COPE model with L predictors. We first define

notation, then we take the second derivative of the likelihood function to calculate the information matrix, and lastly we use the delta method to calculate the variance of $\hat{\gamma}$.

As before, let y_{ijgk} denote the k^{th} outcome in the g^{th} cluster for subject i at time t_{ij} , $k = 1, \dots, k_g$, $g = 1, \dots, G$, $\sum_{g=1}^G k_g = K$, $i = 1, \dots, n$ and $j = 1, \dots, J_i$. Let $Y_{igk} = (y_{i1gk}, \dots, y_{iJ_igk})'$ be the vector of J_i outcomes for subject i and outcome k in cluster g , then the KJ_i -vector of all observations for subject i is $Y_i = (y_{i11}, \dots, y_{i1k_1}, \dots, y_{i1G1}, \dots, y_{iGk_G})'$. Let x_{ijl} be the l^{th} predictor value for subject i at time t_{ij} , $l = 1, \dots, L$, $x_{ij} = (x_{ij1}, \dots, x_{ijL})'$ and define $X_i = (x_{i1}, \dots, x_{iJ_i})'$. Including an intercept for each outcome, we set up the $KJ_i \times G(L+1)$ predictor matrix $M_i = \mathbf{I}_K \otimes (\mathbf{1}, X_i)$, in which $\mathbf{1}$ is a J_i -vector of 1s, $\mathbf{I}_K$ is a $K \times K$ identity matrix and $\otimes$ denotes direct product. We split M_i into $2K$ sets of columns corresponding to the intercepts $\mathbf{1}$ and predictors X_i for each of the K outcomes,

$$M_i = (M_{i1}^{(1,1)}, M_{i2}^{(1,1)}, \dots, M_{i1}^{(1,k_1)}, M_{i2}^{(1,k_1)}, \dots, M_{i1}^{(G,1)}, M_{i2}^{(G,1)}, \dots, M_{i1}^{(G,k_G)}, M_{i2}^{(G,k_G)}).$$

Each $M_{i1}^{(g,k)}$ is a column vector corresponding to the intercept for the k^{th} outcome in cluster g , while $M_{i2}^{(g,k)}$ is an L columned matrix with predictors for the k^{th} outcome in cluster g .

Model (3) can now be written in matrix form as

$$Y_i = M_i \phi + \varepsilon_i. \quad (6)$$

The coefficient vector ϕ is a $G(L+1)$ -vector with elements

$\phi = (\phi'_{11}, \dots, \phi'_{1k_1}, \dots, \phi'_{G1}, \dots, \phi'_{Gk_G})'$, where $\phi_{gk} = (\gamma_{0gk}, \gamma_{1gk} \alpha'_g)'$, γ_{sgk} is the intercept for $s = 0$ and scale parameter for $s = 1$ for the k^{th} outcome in cluster g , $\alpha_g = (\alpha_{g1}, \dots, \alpha_{gL})'$ is an L -vector of fixed effect coefficients for cluster g , α_{gl} is the regression coefficient for the l^{th} predictor in cluster g .

We have K intercepts $\gamma_0 = (\gamma_{011}, \dots, \gamma_{01k_1}, \dots, \gamma_{0G1}, \dots, \gamma_{0Gk_G})'$, K scaling parameters $\gamma_1 = (\gamma_{111}, \dots, \gamma_{11k_1}, \dots, \gamma_{1G1}, \dots, \gamma_{1Gk_G})'$, of which G are fixed to 1 and $K - G$ are to be estimated, and GL coefficients $\alpha = (\alpha'_1, \dots, \alpha'_G)'$. In each cluster we fix the scale parameter of the last outcome in a cluster to 1, $\gamma_{gk_g} \equiv 1$, and let γ_1^* be the $K - G$ sub-vector of γ_1 that omits the G fixed γ s. The covariance matrix of $(\hat{\gamma}_0, \hat{\gamma}_1^*, \hat{\alpha})$ is a $(2K - G + GL) \times (2K - G + GL)$ matrix.

The residual $\varepsilon_i = (\varepsilon_{i11}, \varepsilon_{i12}, \dots, \varepsilon_{i1k_1}, \dots, \varepsilon_{iG1}, \dots, \varepsilon_{iGk_G})'$ is a KJ_i -vector, and we model $\varepsilon_i \sim N(0, V_i(\theta))$, θ is a vector of unknown parameters for the components of the covariance matrix. The log likelihood function of $(\gamma_0, \gamma_1, \alpha, \theta)$ is proportional to

$$-\frac{1}{2} \log |U_i^* V_i U_i^{*'}| - \frac{1}{2} \sum_{i=1}^n (Y_i - M_i \phi)' U_i V_i^{-1} U_i (Y_i - M_i \phi) \quad (7)$$

where U_i is a diagonal matrix with j^{th} diagonal element equal to 1 when y_{ijgk} is observed and 0 when y_{ijgk} is missing and U_i^* is U_i omitting the rows that are all 0. The second derivative of the log likelihood is block diagonal with θ orthogonal to $(\gamma_0, \gamma_1, \alpha)$, so we omit calculations involving θ in the information matrix [25].

We differentiate (7) twice with respect to $(\gamma_0, \gamma_1, \alpha)$. We lay out the estimated information matrix $\hat{I}$ of $(\gamma_0, \gamma_1, \alpha)$ as a 3×3 grid matrix where $\hat{I}_{pq}$ is the second derivative of the log likelihood with respect to p and q , p and q go from 1 to 3, representing γ_0 , γ_1 and α respectively and substituting parameter estimates for the unknown parameters. Because the information matrix is symmetric, we need give only the upper triangle set of blocks

$$\begin{aligned}
& \hat{I}_{11}[(g, k), (g', k')] \\
&= \frac{\partial^2 \ell^2}{\partial \gamma_{0gk} \partial \gamma_{0g'k'}} \\
&= - \sum_{i=1}^n M_{i1}^{(g,k)'} U_i V_i^{-1} U_i M_{i1}^{(g',k')} \hat{I}_{12}[(g, k), (g', k')] \\
&= \frac{\partial^2 \ell^2}{\partial \gamma_{0gk} \partial \gamma_{1g'k'}} \\
&= - \sum_{i=1}^n M_{i1}^{(g,k)'} U_i V_i^{-1} U_i (M_{i2}^{(g',k')} \hat{\alpha}_{g'}) \hat{I}_{13}[(g, k), g'] \\
&= \frac{\partial^2 \ell^2}{\partial \gamma_{0gk} \partial \alpha_{g'}} \\
&= - \sum_{i=1}^n M_{i1}^{(g,k)'} U_i V_i^{-1} U_i \left(\sum_{d=1}^{g'_k} \hat{\gamma}_{1g'd} M_{i2}^{(g',d)} \right) \hat{I}_{22}[(g, k), (g', k')] \\
&= \frac{\partial^2 \ell^2}{\partial \gamma_{1gk} \partial \gamma_{1g'k'}} \\
&= - \sum_{i=1}^n (M_{i2}^{(g,k)} \hat{\alpha}_{g'})' U_i V_i^{-1} U_i (M_{i2}^{(g',k')} \hat{\alpha}_{g'}) \hat{I}_{23}[(g, k), g'] \\
&= \frac{\partial^2 \ell^2}{\partial \gamma_{1gk} \partial \alpha_{g'}} \\
&= - \sum_{i=1}^n (M_{i2}^{(g,k)} \hat{\alpha}_{g'})' U_i V_i^{-1} U_i \left(\sum_{d=1}^{g'_k} \hat{\gamma}_{1g'd} M_{i2}^{(g',d)} \right) \\
&+ \delta(g \\
&= g') \sum_{i=1}^n (Y_i \\
&\quad - M_i \hat{\phi})' U_i V_i^{-1} U_i M_{i2}^{(g,k)},
\end{aligned}$$

and

$$\hat{I}_{33}[g, g'] = \frac{\partial^2 \ell^2}{\partial \alpha_g \partial \alpha_{g'}} = - \sum_{i=1}^n \left(\sum_{k=1}^{g_k} \hat{\gamma}_{1gk} M_{i2}^{(g,k)} \right)' U_i V_i^{-1} U_i \left(\sum_{d=1}^{g'_k} \hat{\gamma}_{1g'd} M_{i2}^{(g',d)} \right),$$

where $V_i = V_i(\theta)|_{\theta=0}$ and $\delta(g = g') = 1$ if $g = g'$, otherwise $\delta(g = g') = 0$. Define $k_g^* = 0$ if $g = 1$ and let $k_g^* = \sum_{i=1}^{g-1} k_i$ for $g \geq 2$ be the cumulative number of outcomes prior to block g , then $\hat{I}_{pq}((g, k), (g', k'))$ is the element in the $(k_g^* + k)^{\text{th}}$ row and $(k_{g'}^* + k')^{\text{th}}$ column of block $\hat{I}_{pq}$ for $q \geq 2$ and $p \geq q$. $\hat{I}_{p3}((g, k), g')$ is the $(k_g^* + k)^{\text{th}}$ row of block $\hat{I}_{p3}$ for $p \geq 2$, and $\hat{I}_{33}(g, g')$ is the g^{th} and g^{th} sub-block of $\hat{I}_{33}$, each sub-block is an $L \times L$ matrix. The covariance matrix of the MLE $(\hat{\gamma}_0, \hat{\gamma}_1^*, \hat{\alpha})$ is the inverse of the negative information matrix deleting the G rows and G columns corresponding to the fixed scaling parameters γ_{1gk_g} , $g = 1, \dots, G$.

We are also interested in inference on $\gamma_{1gk} \hat{\alpha}_g$, $g = 1, \dots, G$ and $k = 1, \dots, k_g - 1$. Define $h \equiv h(\alpha_g, \gamma_{1gk}) = \alpha_g \gamma_{1gk}$, and using the delta method [26, page 45]

$$h(\hat{\alpha}_g, \hat{\gamma}_{1gk}) = \hat{\gamma}_{1gk}^* \hat{\alpha}_g \sim N(\hat{\gamma}_{1gk}^* \alpha_g, \dot{h} \hat{\Psi} \dot{h}') \quad (8)$$

where $\hat{\Psi}$ is the covariance matrix for the parameter subset (γ_{1gk}, α_g) and

$$\dot{h}_{L \times (L+1)} = (\partial h / \partial \gamma_{1gk}, \partial h / \partial \alpha_g) |_{\hat{\alpha}_g, \hat{\gamma}_{1gk}} = (\hat{\alpha}_g, \hat{\gamma}_{1gk} \mathbf{I}_L) \quad (9)$$

Partition

$$\hat{\Psi} = \begin{pmatrix} \hat{\Psi}_{\gamma\gamma} & \hat{\Psi}_{\gamma\alpha} \\ \hat{\Psi}_{\alpha\gamma} & \hat{\Psi}_{\alpha\alpha} \end{pmatrix} \quad (10)$$

Evaluating $\dot{h}$ from (9) in the variance $\dot{h} \hat{\Psi} \dot{h}'$ we get

$$\text{Var}(\hat{\gamma}_{1gk} \hat{\alpha}_{gl}) = \hat{\alpha}_{gl}^2 \hat{\Psi}_{\gamma\gamma} + 2 \hat{\gamma}_{1gk} \hat{\Psi}_{\alpha\gamma, l} \hat{\alpha}_{gl} + \hat{\gamma}_{1gk}^2 \hat{\Psi}_{\alpha\alpha, ll} \quad (11)$$

where $\Psi_{\alpha\alpha, ll}$ is the l^{th} diagonal element of $\Psi_{\alpha\alpha}$ and $\Psi_{\alpha\gamma, l}$ is the l^{th} element of $\Psi_{\alpha\gamma}$.

REFERENCES

1. Weiss, RE. Modeling Longitudinal Data. New York: Springer-Verlag; 2005.
2. Molenberghs, G.; Verbeke, G. Models for Discrete Longitudinal Data. New York: Springer-Verlag; 2005.
3. Diggle, PJ.; Heagerty, PJ.; Liang, KY.; Zeger, SL. Analysis of Longitudinal Data. 2nd Ed.. New York: Oxford University Press; 2002.
4. Davis, CS. Statistical Methods for the Analysis of Repeated Measurements. New York: Springer-Verlag; 2002.
5. McCulloch, CE.; Searle, SR. Generalized, Linear, and Mixed Models. New York: Wiley; 2001.
6. Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag; 2000.
7. Brown, H.; Prescott, R. Applied Mixed Models in Medicine. New York: Wiley; 1999.
8. Littell, RC.; Milliken, GA.; Stroup, WA.; Wolfinger, RD. The SAS System for Mixed Models. Cary, N.C: SAS Institute Inc; 1996.
9. Chaganty NR, Naik DN. Analysis of multivariate longitudinal data using quasi-least squares. Journal of Statistical Planning and Inference. 2002; 103:421–436.
10. Beckett LA, Tancredi DJ, Wilson RS. Multivariate longitudinal models for complex change processes. Statistics in Medicine. 2004; 23:231–239. [PubMed: 14716725]
11. Nummi T, Mottonen J. On the analysis of multivariate growth curve. Metrika. 2000; 52:77–89.
12. Mickey RM, Shema SJ, Vacek PM, Bell DY. Analysis of multiple outcome variables measured longitudinally. Computational Statistics and Data Analysis. 1994; 17:17–33.
13. Dubin JA, Müller HG. Dynamical correlation for multivariate longitudinal data. Journal of the American Statistical Association. 2005; 100(471):872–881.
14. Travison TG, Brookmeyer R. Global effects estimation for multidimensional outcomes. Statistics in Medicine. 2007; 26:4845–4859. [PubMed: 17619238]
15. Gray SM, Brookmeyer R. Estimating a treatment effect from multidimensional longitudinal data. Biometrics. 1998; 54:976–988. [PubMed: 9750246]
16. Gray SM, Brookmeyer R. Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete or time-to-event response variables. Journal of the American Statistical Association. 2000; 95:396–406.

17. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984; 40:1079–1087. [PubMed: 6534410]
18. Sammel M, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. *Journal of Royal Statistical Society, Series B*. 1997; 59:667–678.
19. Lin X, Ryan L, Sammel M, Zhang D, Padungtod C, Xu X. A scaled linear model for multiple outcomes. *Biometrics*. 2000; 56:593–601. [PubMed: 10877322]
20. Rota GC. The number of partitions of a set. *American Mathematical Monthly*. 1964; 71(5):498–504.
21. Young PT. Congruences for Bernoulli, Euler, and Stirling numbers. *Journal of Number Theory*. 1999; 78:204–227.
22. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
23. Chaganty NR. An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference*. 1997; 63:39–54.
24. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. John Wiley & Sons; 2002.
25. Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*. 1989; 84:881–896.
26. Ferguson, TS. *A Course in Large Sample Theory*. Florida: CRC Press; 2002.

Table I

Baseline correlations among the four outcome variables.

	Social Support	Positive Action	Anxiety	Somatization
Social support	1.00	.49	.25	.24
Positive action	.49	1.00	.19	.17
Anxiety	.25	.19	1.00	.75
Somatization	.24	.17	.75	1.00

Table II

Summary statistics for all 15 COPE models for four outcomes, ordered by Akaike's Information Criterion (AIC) from best to worst. Model identifies the clustering among outcomes; Outcomes 1–4 represent Social Support= 1, Positive Action= 2, Anxiety= 3 and Somatization= 4. AIC, Bayes Information Criterion (BIC) and $-2 \log L$ are in smaller-is-better form. Diff is the $-2 \log$ likelihood ratio test (LRT) statistic between each model and the saturated model 1, 2, 3, 4; df is the degrees of freedom for the test and the p -value comes from comparing the LRT statistic to the $\chi^2(df)$ reference distribution. Columns labeled γ_1 to γ_4 are the estimates of the scale γ_1 s from the link function; the γ_1 s correspond to the outcomes in the order shown in the Model column.

Order	Model	AIC	BIC	$-2 \log L$	Diff	df	p -value	γ_1	γ_2	γ_3	γ_4
1	134, 2	12,325.5	12,477.3	12,245.5	23.7	18	.17	.38	.88	1.00	1.00
2	1, 2, 34	12,330.6	12,520.4	12,230.6	8.8	9	.46	1.00	1.00	.87	1.00
3	13, 2, 4	12,333.7	12,523.5	12,233.7	11.9	9	.22	.48	1.00	1.00	1.00
4	14, 2, 3	12,337.8	12,527.6	12,237.8	16.0	9	.07	.37	1.00	1.00	1.00
5	1, 234	12,339.2	12,491.0	12,259.2	37.4	18	.00	1.00	.38	.90	1.00
6	1234	12,339.3	12,453.2	12,279.3	57.5	27	.00	.39	.23	.88	1.00
7	1, 2, 3, 4	12,341.8	12,569.6	12,221.8	-	-	-	1.00	1.00	1.00	1.00
8	13, 24	12,344.9	12,496.8	12,264.9	43.1	18	.00	.53	1.00	.35	1.00
9	12, 34	12,345.1	12,497.0	12,265.1	43.3	18	.00	1.05	1.00	.88	1.00
10	14, 23	12,346.8	12,498.6	12,266.8	45.0	18	.00	.51	1.00	.59	1.00
11	123, 4	12,347.8	12,499.7	12,267.8	46.0	18	.00	.47	.28	1.00	1.00
12	1, 23, 4	12,348.4	12,538.2	12,248.4	26.6	9	.00	1.00	.53	1.00	1.00
13	1, 24, 3	12,350.8	12,540.6	12,250.8	29.0	9	.00	1.00	.37	1.00	1.00
14	124, 3	12,351.6	12,503.4	12,271.6	49.8	18	.00	.38	.23	1.00	1.00
15	12, 3, 4	12,356.4	12,546.2	12,256.4	34.6	9	.00	1.06	1.00	1.00	1.00

Table III

Comparison of coefficient estimates, standard errors and *p*-values from the saturated model (Sat) and for $\alpha_g * \gamma_{1gk}$ from the COPE model 134.2, where Social Support= 1, Anxiety= 3 and Somatization= 4 are in the same cluster, and Positive Action= 2 is in a separate cluster.

Predictor	Social Support		Positive Action		Anxiety		Somatization	
	Sat	COPE	Sat	COPE	Sat	COPE	Sat	COPE
Point Estimate								
Month	-.010	-.010	-.016	-.016	-.026	-.023	-.024	-.026
Female	.171	.163	.036	.034	.348	.379	.444	.429
Age	.005	.002	.042	.040	.011	.006	-.005	.006
Spring	-.031	.024	-.080	-.050	.054	.055	.096	.063
Summer	-.103	-.029	-.092	-.053	-.052	-.066	-.039	-.075
Month18	.016	.013	.024	.023	.031	.030	.030	.034
Month36	-.014	-.005	-.014	-.010	-.010	-.011	-.002	-.012
Drug NUU	.063	.027	.073	.053	.007	.064	.052	.072
Drug UU	.033	.086	.092	.117	.223	.198	.223	.225
Parental Alcohol	-.036	-.051	-.026	-.036	-.082	-.119	-.158	-.135
Std Error								
Month	.004	.003	.005	.004	.005	.005	.005	.005
Female	.044	.032	.055	.040	.093	.075	.099	.086
Age	.004	.007	.005	.009	.007	.015	.007	.017
Spring	.036	.017	.042	.038	.053	.040	.051	.045
Summer	.038	.019	.044	.040	.053	.040	.051	.045
Month18	.006	.004	.007	.006	.008	.007	.008	.007
Month36	.005	.003	.006	.006	.009	.007	.009	.007
Drug NUU	.047	.028	.058	.041	.096	.064	.102	.073
Drug UU	.066	.036	.080	.062	.112	.080	.116	.089
Parental Alcohol	.040	.022	.047	.039	.063	.047	.062	.054
P-Value								
Month	.014	.000	.001	.000	.000	.000	.000	.000
Female	.000	.000	.517	.398	.000	.000	.000	.000
Age	.196	.717	.000	.000	.097	.713	.464	.710
Spring	.396	.167	.059	.184	.309	.166	.058	.168
Summer	.007	.133	.037	.179	.330	.100	.446	.096
Month18	.007	.001	.000	.000	.000	.000	.000	.000

Predictor	Social Support		Positive Action		Anxiety		Somatization	
	Sat	COPE	Sat	COPE	Sat	COPE	Sat	COPE
Month36	.007	.145	.016	.082	.265	.108	.803	.100
Drug NUU	.177	.335	.204	.198	.938	.320	.608	.321
Drug UU	.620	.018	.250	.061	.046	.013	.054	.012
Parental Alcohol	.359	.018	.573	.353	.196	.012	.011	.014