Joint and Conditional Models for Repeated Measures Data with a Time Varying Covariate

Charlie Xiang Zhang
IDEC Pharmaceuticals
3030 Callan Road
San Diego, CA 92121

Wenzhi Li
Department of Statistics
Stanford University
Stanford, CA 94305
wenzhi@statistics.stanford.edu

Robert E. Weiss*
Department of Biostatistics
UCLA School of Public Health
Los Angeles, CA 90095-1772
rob@rem.ph.ucla.edu

June 2, 2000

Abstract

Let Y_i and W_i be n–vectors of repeated measures on subject i where respective elements of Y_i and W_i are measured contemporaneously. In practice, the most common repeated measures approach to studying the conditional distribution of $Y_i|W_i$ is to include W_i as a single covariate in the univariate repeated measures model $Y_i = X_i\alpha_{p\times 1} + W_i\gamma + \delta_i, \delta_i \sim N_n(0, \Sigma(\theta))$, where γ is a scalar. This model permits a very restricted mechanism for the way W_i affects Y_i. We generalize this model to permit every element of W_i to affect Y_i by allowing $Y_i = X_i\alpha + MW_i + \delta_i$, with M an $n \times n$ matrix of unknown parameters. To guide in the specification of M, we consider the joint

*Contact author.

multivariate modeling of Y_i and W_i under various parameterized covariance structures for the error variance. The joint and conditional modeling approaches are compared.

Key Words: Autoregressive Structure, Bivariate Repeated Measure, Brownian Motion, Covariance Structure, Product Normal Covariance, Random Effects Model.

1 Introduction

Many authors (Laird and Ware, 1982; Jennrich and Schluchter, 1986; Verbyla and Cullis, 1990) have explored repeated measures models for continuous response data. In this paper, we consider the situation where we have two time-varying measurements Y_i and W_i. In most analyses, one repeated measure $Y_i, n \times 1$ is treated as the dependent variable, and its relation to W_i is explored by treating W_i and X_i equivalently as covariates. The covariance structure for Y_i is usually specified as one of several standard forms, such as Simple, Compound symmetry, Autoregressive(1), Independent Increments, Unstructured, and so forth. These models can be fit with SAS Proc Mixed (SAS Institute, 1996) or BMDP5V (Dixon, 1992). While multiple repeated measures are common, research that accounts for the multivariate nature of the repeated observations is rare. Exceptions include Reinsel (1984), and Shah, Laird and Schoenfeld (1997) who use a multivariate random effects models and Sy, Taylor and Cumberland (1997) who use a stochastic process to model (Y_i, W_i). These papers are mainly concerned with computation, although Sy, Taylor and Cumberland (1997) and Reinsel (1984) concern themselves slightly with the conditional distribution of $Y_i|W_i$.

Suppose that $Y_i = (Y_{i1}, \ldots, Y_{in})'$ and $W_i = (W_{i1}, \ldots, W_{in})'$ are two repeated measures on subject i, with (Y_{ij}, W_{ij}) a bivariate measurement at time t_j where $j > j'$ implies $t_j > t_{j'}$. For ease of both presentation and analysis, we assume n is the number of repeated measures for all variables for all N subjects, and matrix X_i of order $n \times p$ contains known covariates.

The usual modeling approach directly fits a conditional model $Y_i|W_i$, in which Y_i is the dependent variable, and W_i is treated as a covariate.

$$\begin{aligned} Y_i &= X_i\alpha + W_i\gamma + \delta_i \\ \delta_i &\sim N_n(0, \Sigma(\theta)) \end{aligned} \tag{1}$$

where $\alpha, p \times 1$ is a vector of unknown regression parameters, θ is a vector of unknown covariance structure parameters, and γ is a scalar regression parameter of particular interest. There are three qualitative outcomes in this model: γ can be negative, zero or positive, with obvious interpretations.

This model can be generalized to allow for the proper potential effect of W_i on Y_i. Consider the more general model

$$Y_i = X_i\alpha + MW_i + \delta_i\,, \tag{2}$$

where $M = (m_{jk}), n \times n$ is a matrix of unknown regression parameters. Model (1) is a special case of (2) where $M = \gamma I$ is a scalar matrix. Many other important special cases of M are possible. (i) The matrix M could be diagonal, $M = \text{diag}(m_{11}, \ldots, m_{nn})$. This can be fit using standard software with covariates $\text{diag}(W_i)$ appended to the X_i matrix. This implies that the W_{ij} measured at the same time as Y_{ij} affects Y_{ij} but the strength of the effect changes over time. This is a time-varying coefficient model in contrast to (1) which has a time-fixed coefficient. (ii) One could have M lower diagonal. To predict Y_{ij}, we need to know $W_{i1}, \ldots, W_{ij}$, but future observations $W_{ij'}, j' > j$ do not help in the prediction. (iii) We could have $M = (m_1 - m_2)I + m_2\mathbf{11'}$ which we call compound symmetry (CSYM); if $m_1 > m_2 > 0$, this suggests a positive relation between Y_i and W_i, with the contemporaneously measured W_{ij} the most important in predicting Y_{ij}, but other $W_{ij'}, j' \neq j$ also contribute to the prediction. The contribution of the observations $W_{ij'}, j' \neq j$ is equal. (iv) The matrix M could be unstructured. (v) Other structures for M could be appropriate; for example, a special case of the independent increments model considered later has an M with non-zero elements only on the diagonal and first column.

A second approach to modeling assumes Y_i and W_i have a joint distribution with X_i treated as covariates. From this joint distribution, we can deduce the conditional distribution $Y_i|W_i$. For our joint model, we assume

$$\begin{pmatrix} Y_i \\ W_i \end{pmatrix} = \begin{pmatrix} X_i\alpha_1 \\ X_i\alpha_2 \end{pmatrix} + \epsilon_i, \tag{3}$$

$$\epsilon_i \sim N(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is a $2n \times 2n$ symmetric positive definite matrix. Here $\Sigma = \Sigma(\theta)$ is a function of a parameter vector $\theta = (\theta_1, \ldots, \theta_q)'$. Based on (3), we have the conditional distribution of $Y_i|W_i$ which produces the model

$$
\begin{aligned}
Y_i &= X_i\alpha_1 + \Sigma_{12}\Sigma_{22}^{-1}(W_i - X_i\alpha_2) + \delta_i, \qquad (4)\\
\delta_i &\sim N\left(0, \Sigma_{11.2}\right)
\end{aligned}
$$

where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. We are interested in the coefficients $M = \Sigma_{12}\Sigma_{22}^{-1}$ and the covariance matrix $\Sigma_{11.2}$. By fitting the joint model, we get information about the relation between Y_i and W_i. For example, does Y_{ij} increase when W_{ij} increases or does Y_{ij} increase when $W_{ij'}$, $j \neq j'$ increases?

Our main interest is in models which can be fit using standard linear software. When $MX_i\alpha_2$ is nonlinear in the parameters, that suggests that the user take the joint modeling approach rather than the conditional approach. An important variant of (4) is model (2), obtained by omitting $-\Sigma_{12}\Sigma_{22}^{-1}X_i\alpha_2$ in (4). This model can be fit using standard software. Essentially this ignores the complexities in modeling the effects of X_i that are caused by conditioning on W_i. When it is easy to incorporate the $MX_i\alpha_2$ term into model (4), we comment on how to do. Easy here is defined as a linear model which is a superset of the correct model. The advantage of the conditional model is that it does not require modeling of the marginal distribution of W_i. The advantage of the joint model is that the model becomes linear.

The next section analyzes the consequences of various specifications for Σ in model (3). We see under what circumstances we recover model (1) or

(2). We apply some of the methods in section 3.

2 Analysis of Models with Parameterized Covariance Matrices

We calculate M and $\Sigma_{11.2}$ assuming several important models for Σ.

2.1 Product Normal Model

The covariance matrix for the product normal model is in the form

$$\Sigma = D\left(R \bigotimes \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) D, \tag{5}$$

where D is a $2n \times 2n$ diagonal matrix with the standard deviations of Y_{ij} and then W_{ij} on the diagonal; R is a correlation matrix, that is, R is positive definite with diagonal entries 1. In partitioned form

$$\begin{aligned} \Sigma &= \begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix} \begin{pmatrix} R & \rho R \\ \rho R & R \end{pmatrix} \begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix} \\ &= \begin{pmatrix} D_1 R D_1 & \rho D_1 R D_2 \\ \rho D_2 R D_1 & D_2 R D_2 \end{pmatrix}. \end{aligned} \tag{6}$$

In (6) and in the future, blanks indicate zeros. In this model Y_{ij} and W_{ij} have a constant correlation ρ, independent of subject i and time j, and Y_i and W_i have the same correlation matrix R. Straightforward calculation gives

$$M = \Sigma_{12}\Sigma_{22}^{-1} = \rho D_1 R D_2 (D_2 R D_2)^{-1} = \rho D_1 D_2^{-1}.$$

This is diagonal and it implies that Y_{ij} is only affected by W_{ij}, not other $W_{ik}, k \neq j$.

Homogeneity of variances for Y_i and of W_i is an important special case. Let $D_1 = d_1 I_n$ and $D_2 = d_2 I_n$, then $\rho D_1 D_2^{-1} = \rho d_1 d_2^{-1} I_n$ is a scalar times the identity or a scalar matrix. When we fit the conditional model, we simply add one more covariate, i.e., W_i, with coefficient γ. Model (1) is actually correct, and the new α would be equal to $\alpha_1 - \rho d_1 d_2^{-1} \alpha_2$ under the joint model. This also holds if $D_1 D_2^{-1}$ is a scalar matrix. If $\rho D_1 D_2^{-1}$ is not γI, then complete execution of the conditional paradigm requires that we fit a mean structure not of $\mu_i = X_i \alpha$, but rather of $\mu_i = D_1 D_2^{-1} X_i \alpha \rho$. Now we could either fit (2), or to avoid combining mean function and variance function parameters, we could fit a more general model with mean $\mu_i = \Psi X_i \alpha$, where Ψ is a diagonal matrix of unknown parameters. Since there is now one free parameter, we might either fix the first element of Ψ equal to one, or fix the trace of Ψ^2 to be one. Our preference at this point is to fit the bivariate model rather than a correct conditional model that is non-linear in the parameters.

Whether D_1 and D_2 are scalar or diagonal, the conditional covariance matrix of Y_i

$$\Sigma_{11.2} = (1 - \rho^2) D_1 R D_1$$

is independent of D_2 and it is reduced by a factor $(1 - \rho^2)$ over the marginal covariance of Y_i.

2.2 Random Intercept Model

In the random intercept model

$$\Sigma_{jk} = \sigma_{jk}(I_n + d_{jk} 1_n 1_n'), \quad j, k = 1, 2 \tag{7}$$

that is,

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \sigma_{11} I_n & \sigma_{12} I_n \\ \sigma_{12} I_n & \sigma_{22} I_n \end{pmatrix} + \begin{pmatrix} 1_n & \\ & 1_n \end{pmatrix} \begin{pmatrix} \sigma_{11} d_{11} & \sigma_{12} d_{12} \\ \sigma_{12} d_{12} & \sigma_{22} d_{22} \end{pmatrix} \begin{pmatrix} 1_n & \\ & 1_n \end{pmatrix}' \\
&= \begin{pmatrix} \sigma_{11}(I_n + d_{11} 1_n 1_n') & \sigma_{12}(I_n + d_{12} 1_n 1_n') \\ \sigma_{12}(I_n + d_{12} 1_n 1_n') & \sigma_{22}(I_n + d_{22} 1_n 1_n') \end{pmatrix}
\end{aligned} \tag{8}$$

where 1_n is a column vector of 1's and blanks in the arrays indicate zeros. This model is equivalent to the model

$$\begin{pmatrix} Y_i \\ W_i \end{pmatrix} = \begin{pmatrix} X_i \alpha_1 \\ X_i \alpha_2 \end{pmatrix} + \begin{pmatrix} 1_n & \\ & 1_n \end{pmatrix} B_i + \delta_i^{\dagger}, \tag{9}$$

where B_i and $\delta_i^{\dagger}$ are independent of each other. The random intercept B_i is bivariate normal with variance

$$\mathrm{Cov}(B_i) = \begin{pmatrix} \sigma_{11} d_{11} & \sigma_{12} d_{12} \\ \sigma_{12} d_{12} & \sigma_{22} d_{22} \end{pmatrix},$$

The error term $\delta_i^{\dagger}$ is multivariate normal, mean zero and covariance

$$\mathrm{Cov}(\delta_i^{*}) = \begin{pmatrix} \sigma_{11} I_n & \sigma_{12} I_n \\ \sigma_{12} I_n & \sigma_{22} I_n \end{pmatrix},$$

Model (3) with covariance (7) is equivalent to model (9) after integrating out B_i. We have

$$\begin{aligned}
M &= \sigma_{12}(I_n + d_{12} 1_n 1_n') \sigma_{22}^{-1} (I_n + d_{22} 1_n 1_n')^{-1} \\
&= \frac{\sigma_{12}}{\sigma_{22}} \left(I_n + \frac{d_{12} - d_{22}}{1 + n d_{22}} 1_n 1_n' \right)
\end{aligned} \tag{10}$$

and the covariance matrix

$$\Sigma_{11.2} = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} I_n + \left[\sigma_{11} d_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} \left\{ d_{12} + (d_{12} - d_{22}) \frac{1 + n d_{12}}{1 + n d_{22}} \right\} \right] 1_n 1_n'$$

As (10) has the form $\Sigma_{12}\Sigma_{22}^{-1} = aI_n + b1_n1_n'$, implicitly defining a and b, we have the conditional model

$$\begin{aligned} Y_i &= X_i\alpha_1 + (aI_n + b1_n1_n')(W_i - X_i\alpha_2) + \epsilon_i \\ &= aW_i + b1_n1_n'W_i + X_i(\alpha_1 - a\alpha_2) - 1_n1_n'X_i(b\alpha_2) + \epsilon_i \end{aligned}$$

We can reparameterize this model by

$$\begin{aligned} \beta_1 &= \alpha_1 - a\alpha_2 \\ \beta_2 &= -b\alpha_2\,. \end{aligned}$$

The conditional model becomes linear in the parameters and in covariates W_i and X_i,

$$Y_i = aW_i + b1_n1_n'W_i + X_i\beta_1 + 1_n1_n'X_i\beta_2 + \epsilon_i$$

By (4) we can see that, in the conditional random intercept model, correct specification of the covariate matrix X_i should also include predictors $1_n1_n'W_i$ and $1_n1_n'X_i$. Since the conditional model still has random effects structure, we can use regular software to fit the conditional model. An interesting characteristic of the definition of b is that it depends on n, the number of repeated measures. Thus if data (Y_{ij}, W_{ij}) is missing in the design, we cannot fit the usual conditional random effects model, but rather must use some manner of imputation scheme to create a balanced data set. This would require modeling the marginal distribution of W_i, something we hoped to avoid by using the conditional approach.

Suppose that we have a general multivariate random effects structure with

$$\Sigma_{jk} = \sigma_{jk}(I_n + Z_jD_{jk}Z_k'), \quad j,k = 1,2 \tag{11}$$

which comes from the bivariate random effects model with $Y_i = X_i\alpha + Z_1\beta_{1i} + \delta_{1i}$ and $W_i = X_i\alpha + Z_2\beta_{2i} + \delta_{2i}$ with $\mathrm{Cov}(\beta_{ji}, \beta_{ki}) = \sigma_{jk}D_{jk}$, $\mathrm{Cov}(\delta_{ji}, \delta_{ki}) = I\sigma_{jk}$, and with the β_{ji}'s $j = 1, 2$ independent of the δ_{ki}'s, $k = 1, 2$. As with the random intercept model already considered, we need equality of the Z_1 and Z_2 matrices across cases, although we permit potentially different random effects for W and for Y. Then the conditional model is also a random effects model,

$$Y_i = X_i\alpha^* + W_i\gamma + Z^*\Omega_X Z_2' X_i + Z^*\Omega_W Z_2' W_i + Z^* B_i^* + \delta_i^* \qquad (12)$$

where the fixed effects include X_i, W_i and also growth curve like terms $Z^*\Omega_W Z_2' W_i$ and $Z^*\Omega_X Z_2' X_i$, where Z^* is a matrix that spans the column space of (Z_1, Z_2); if $Z_1 = Z_2$, then $Z^* = Z_1$; and Ω_W and Ω_X are regression parameter matrices of unknown coefficients. The random effects portion should include random effects for $Z^* B_i^*$ with unstructured covariance matrix for the B_i^*'s and the usual normal mean zero error term δ_i^* with covariance matrix a constant $\sigma^{2*}I$. If $Z_1 = Z_2 = 1_n$, then this reduces to the case already considered. The exact relationship of the parameters in this conditional model are complicated functions of the parameters of the joint model and we omit the unenlightening algebra. In practice in (12), we would take all regression parameters α^*, γ, Ω_X, and Ω_W as parameters to be estimated.

2.3 Multivariate AutoRegressive Structure

We denote by X'_{ij} the jth row of the design matrix X_i for subject i. The Multivariate AutoRegressive-1 (MVAR(1)) model is defined by

$$\begin{aligned}
\begin{pmatrix} Y_{i1} - X'_{i1}\alpha_1 \\ W_{i1} - X'_{i1}\alpha_2 \end{pmatrix} &= \epsilon_{i1}, \\
\epsilon_{i1} &\sim N(0, \Sigma_0) \\
\begin{pmatrix} Y_{ij} - X'_{ij}\alpha_1 \\ W_{ij} - X'_{ij}\alpha_2 \end{pmatrix} &= A \begin{pmatrix} Y_{ij-1} - X'_{ij-1}\alpha_1 \\ W_{ij-1} - X'_{ij-1}\alpha_2 \end{pmatrix} + \epsilon_{ij}, \\
\epsilon_{ij} &\sim N(0, V)\,, \quad j = 2, 3, \ldots, n\,.
\end{aligned}$$

Both V and Σ_0 are 2×2 covariance matrices. We restrict our analytic results to $A = \text{diag}(\mu, \lambda)$, 2×2, although A could be a full matrix with eigenvalues between zero and one which we consider in our example. Under stationarity, we have $V = \Sigma_0 - A\Sigma_0 A'$ and

$$\text{Cov} \begin{pmatrix} Y_{ij} - X'_{ij}\alpha_1 \\ W_{ij} - X'_{ij}\alpha_2 \end{pmatrix} = \Sigma_0.$$

It is convenient to order the observations of the ith subject as $Y_{i1}, W_{i1}, \ldots, Y_{in}, W_{in}$ to display the covariance matrix.

$$\text{Cov} \begin{pmatrix} Y_{i1} \\ W_{i1} \\ \ldots \\ Y_{in} \\ W_{in} \end{pmatrix} = \begin{pmatrix} \Sigma_0 & \Sigma_0 A' & \Sigma_0 A^{2'} & \ldots & \Sigma_0 A^{(n-1)'} \\ A\Sigma_0 & \Sigma_0 & \Sigma_0 A' & \ldots & \Sigma_0 A^{(n-2)'} \\ A^2\Sigma_0 & A\Sigma_0 & \Sigma_0 & \ldots & \Sigma_0 A^{(n-3)'} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ A^{(n-1)}\Sigma_0 & A^{(n-2)}\Sigma_0 & A^{(n-3)}\Sigma_0 & \ldots & \Sigma_0 \end{pmatrix}_{2n \times 2n}, \tag{13}$$

where $A^{n'}$ denotes the transpose of A^n. However, $\Sigma_{11}, \Sigma_{12}, \text{and} \Sigma_{22}$ are still defined as after equation (3). The matrix M is given by

$$\frac{1}{1 - \lambda^2} \frac{\sigma_{12}}{\sigma_{22}} *$$

$$\begin{pmatrix} 1-\lambda^2 & & & & & \\ \mu-\lambda & 1-\mu\lambda & & & & \\ \mu(\mu-\lambda) & (\mu-\lambda)(1-\mu\lambda) & 1-\mu\lambda & & & \\ \vdots & \vdots & \vdots & \cdots & \ddots & \\ \mu^{n-2}(\mu-\lambda) & \mu^{n-3}(\mu-\lambda)(1-\mu\lambda) & \mu^{n-4}(\mu-\lambda)(1-\mu\lambda) & \cdots & (\mu-\lambda)(1-\mu\lambda) & 1-\mu\lambda \end{pmatrix},$$

a lower triangular matrix that is a function of λ, μ, and $\sigma_{12}\sigma_{22}^{-1}$. This implies that in this model, the current observations are only affected by current and past observations, not future observations. If μ and λ are positive and less than one, then, ignoring the constant out front, the entries in the matrix are less than 1 in absolute value and largest along the diagonal. Except for the first column and the first subdiagonal, coefficients decrease geometrically below and to the left of the diagonal. In fitting MW in our example, we take M to be a lower diagonal matrix of general form rather than the nonlinear parameterization given above.

2.4 Independent Increments and Brownian Motion

The independent increments (II) (Louis 1988) model is another model where M is lower triangular. In this model $\epsilon_{i1} \sim N(0, A_1)$, and

$$\mathrm{Cov}(\epsilon_{ij}) = \mathrm{Cov}\begin{pmatrix} Y_{ij} - X'_{ij}\alpha_1 \\ W_{ij} - X'_{ij}\alpha_2 \end{pmatrix} = A_j.$$

And the covariance $\mathrm{Cov}(\epsilon_{ij}, \epsilon_{ij'}) = A_{\min(j,j')}$. This model comes from the idea that $\epsilon_{ij} = \epsilon_{i(j-1)} + \delta_{ij}$ with the $\delta_{ij} \sim N(0, A_j - A_{j-1})$ all independent.

Let the elements of A_j be

$$A_j = \begin{pmatrix} a_j & b_j \\ b_j & c_j \end{pmatrix}$$

The submatrices Σ_{11}, Σ_{12}, and Σ_{22} all have the II form with elements a_j, b_j, and c_j respectively, where the jk^{th} element of each matrix is $a_{\min(j,k)}$, $b_{\min(j,k)}$, or $c_{\min(j,k)}$. Then $M =$

$$\begin{pmatrix} \frac{b_1}{c_1} & & & & \\ \frac{b_1}{c_1} - \frac{b_2-b_1}{c_2-c_1} & \frac{b_2-b_1}{c_2-c_1} & & & \\ \frac{b_1}{c_1} - \frac{b_2-b_1}{c_2-c_1} & \frac{b_2-b_1}{c_2-c_1} - \frac{b_3-b_2}{c_3-c_2} & \frac{b_3-b_2}{c_3-c_2} & & \\ \vdots & & & \ddots & \\ \frac{b_1}{c_1} - \frac{b_2-b_1}{c_2-c_1} & \frac{b_2-b_1}{c_2-c_1} - \frac{b_3-b_2}{c_3-c_2} & \frac{b_3-b_2}{c_3-c_2} - \frac{b_4-b_3}{c_4-c_3} & \cdots & \frac{b_n-b_{n-1}}{c_n-c_{n-1}} \end{pmatrix}$$

is lower triangular. The conditional covariance matrix has the II form.

Brownian motion (for example Sy et al. 1997) is a special case of the II model. Two versions of Brownian motion are (a) the constant variance increments where $A_j = jA_0$, with A_0 positive definite, when $M = \Sigma_{21}\Sigma_{22}^{-1} = b_1 c_1^{-1} I$ is a scalar matrix, and (b) $A_j = A_0 + (j-1)B$ with B also positive definite, when $MW_i = \gamma_1 W_i + \gamma_2 W_{i1} 1_n$, where γ_1 and γ_2 are unknown parameters to be estimated. In the first form, model (1) is appropriate. For the second form, M has a non-zero first column and non-zero diagonal matrix, and we need to include W_{i1} as a time-fixed covariate, and with this modification, (1) is still appropriate.

3 Data: Pain Ratings and Tolerance

The data we use are from a pain tolerance experiment (Fanurik, et al. 1993) carried out at UCLA. Weiss, Wang and Ibrahim (1997) analyzed log pain tolerance for the 58 children with complete data and developed a predictive Bayesian approach to model selection in the random effects model. Pain tolerance is the number of seconds that a child can keep his/her hand immersed in very cold water. The $X_{i\ 4\times 8}$ matrix for each child has an intercept, an indicator for child's coping style (CS), either 0 for attender (A) or 1 for distractor (D); and 6 columns of indicators for CS by treatment (TMT) interaction. Treatment may be counseling to attend (A), distract (D) or null (N) and we assume it interacts with CS, to produce 6 possible CS*TMT groups, labeled AA, AD, AN, DA, DD, and DN. Since TMT was applied between the third and fourth trials, the last 6 columns are all zeros except for a single 1 in row 4 indicating which of the 6 groups the child belongs to. This portion of the data is the same as that used in Weiss et al. (1997). For each child at each trial, we also have a pain rating on a 1 (least painful) to 10 (most painful) scale. We wish to assess the effect of pain rating W_i on log pain tolerance Y_i; do unusually high pain ratings account for any decrease in pain tolerance ability? Also we will assess whether the random effects covariance structure is the most appropriate one for these data.

We fit models (2) and (3) to this data, omitting the II structure. The II structure is inappropriate since the marginal variance is not increasing over time. Table 1 gives the meanings of the α_{jk} parameters in the joint model. Table 2 gives a summary of results for the joint model. It includes the names of the models, the number of parameters in the covariance structure, the maximized log likelihoods, AIC penalized on the number of covariance parameters ignoring the 16 parameters in α, and the log likelihood ratio test statistics, degrees of freedom (df) and p-values for testing against the unstructured model. The last three columns give p-values for tests of $H_0 : \alpha_j = 0$ for three parameters, α_{12}, α_{17}, and α_{24}. These three parameters and α_{26} were the only regression coefficients significantly different from zero. The p-value for α_{26} was always smaller than .0001. P-values (and estimates and standard errors, not shown) vary with model. One anomalous p-value is apparent for the random intercept model and α_{24}. These results imply that distractors have longer baseline pain tolerance than attenders, that the distract treatment improves pain tolerance for distractors, that the distract treatment for attenders and the attend treatment for distractors increase pain ratings. Treatments mis-matched with child CS increase the perceived pain. The conclusions (significant / not significant at the .05 level) regarding α_{24} depend on the choice of model.

Model selection is performed using likelihood ratio tests. The unstructured model is significant compared to either of the two MVAR(1) models or to the random intercept model. The two product normal models are not rejected when compared to the unstructured model, and the equal variance model is not rejected when compared to the unequal variance model. AIC also supports this conclusion. This is very convenient, because it means that pain rating can be used as a single column predictor in the X_i matrix of the conditional model. It implies that pain ratings are associated with the contemporaneously measured log pain tolerance, but do not affect the other times. The estimate and (standard error) of the diagonal element of $M = \Sigma_{12}\Sigma_{22}^{-1}$ for the equal variance product normal model is $-.046(.019)$.

The unequal variance product normal model estimates vary from this by at most .005. The estimates from the random intercept model are $-.048(.02)$ on the diagonal and $-.007(.014)$ off diagonal. The coefficients for the MVAR(1) diagonal A model are negative, and have structural zeros for the elements above the diagonal. The unstructured A model has negative coefficients down and below the diagonal, and all positive coefficients above the diagonal, but the standard errors are all large compared to the coefficients except along the diagonal. Estimates from the unstructured model differ rather dramatically from estimates from the other models, but the standard errors are also correspondingly quite high.

We can also compare these joint models against the nested models of independence of Y_i and W_i. The product normal models reject with a chisquare of 6 on 1 df for a p-value less than .02. The random intercept model also has a test statistic of 5.9, but is on 2 df, and has a p-value of .052. Inspection suggests that the random effects are not correlated, and most of the increase in log likelihood is due to correlation among the δ_i errors of equation (9). The MVAR(1) with diagonal A very significantly rejects the hypothesis of independence. Only the unstructured model is unable to reject the hypothesis of independence, probably due to its excess of parameters and corresponding high standard errors.

Table 3 gives results of fitting the conditional model (2). Definitions of the entries are given in the caption; all tests are likelihood ratio tests. We fit 5 models, conditional versions of all the models in table 2 except for the

MVAR(1), full A model. Model C1 is the conditional version of the product normal model, which gives a diagonal M and unstructured correlation matrix with constant variance. Model C2 comes from the product normal model with non-constant variances. Model C3 is the conditional version of the random intercept model; CSYM is compound symmetry where residual covariance structure is that of a random intercept model. Model C4 has an unstructured lower triangular M, the conditional version of the MVAR(1) with diagonal A matrix. We fit an unstructured M and covariance matrix rather than attempt a complicated non-linear parameterized covariance structure as would happen for the MVAR(1) with general A.

None of the four simpler models can be rejected at the .05 level when tested against the unstructured model, although model C3 comes closest to being rejected. Model C1 is also nested in model C2, and is not rejected when the two are compared, and C2 is nested in C4, and again cannot be rejected. Model C1 is the second most parsimonious model and has the highest AIC, essentially equal to AIC from model C2. In testing $H_0 : M = 0$, the product normal models reject the null hypothesis, and the compound symmetry model does not quite reject at .05. The conditional version of the MVAR(1) model has many more parameters in M than the joint model and like the unstructured model, is unable to reject H_0. If the joint likelihood factored into conditional and marginal likelihoods with separate parameters $L(Y, W|\nu, \phi) = L(Y|W, \nu)L(W|\phi)$, we would expect the likelihood ratio test of $M = 0$ to be identical in the joint and conditional models. For the product normal models, the likelihood does not factor, and the conditional model was too complex for the MVAR(1) model. For CSYM and the unstructured models, the tests are the same, up to slightly different estimates of the likelihood possibly resulting from using different packages for the two calculations. We have a preference for model C1, partly based on the statistical results, partly based on parsimony, and strongly influenced by the results from fitting the joint model, where the preference of the data for the product normal, constant variance model is strong.

For model C1, M is scalar with estimate (se) $-.039(.018)$. For the diagonal M, diagonal estimates are $(-.033(.018), -.050(.02), -.059(.020), -.047(.030))$. The compound symmetry model has diagonal element $-.048(.020)$ and off diagonal element $-.077(.014)$. These results are qualitatively similar to the estimates from the joint model. When M is unstructured, standard errors are high and any pattern in the results is difficult to see.

4 Discussion

The choice between the joint and conditional modeling approaches can be made depending on which structure provides an easier mechanism for setting up the model; that is, one should choose the approach which makes it easier to think about the relationship between Y_i and W_i. We found the joint approach easier to use. The issues in specifying the M matrix in the conditional approach were not apparent to us until we first considered the joint model. The joint model also gave simpler parameterized structures in our examples. For example for the MVAR(1) model with diagonal A, we used fewer parameters in fitting the covariance structure using the joint approach. Of course, if we were to fit the exact conditional distribution we would have fewer parameters since the marginal of W_i need not be modeled, but this would have required fitting complicated nonlinear models. The joint approach also appropriately partials out the information in X_i for modeling W_i, while this is difficult in the conditional approach – we basically ignored the term $-\Sigma_{12}\Sigma_{22}^{-1}X_i\alpha_2$ in our conditional approach (2). The joint approach also supplies more information, since it includes information about the marginal distribution of W_i, as well as the marginal distribution of Y_i and both conditional distributions $W_i|Y_i$ and $Y_i|W_i$. We did not consider the unbalanced case in this paper. The conditional approach can be quite difficult when dealing with unbalanced data. Depending on the structure assumed, the M matrix will be different for cases with different numbers of observations and spacings of repeated measures. The random intercept model and the general form of the independent increments model are both examples with this difficulty. Imputation of the missing data is required before fitting the conditional models, which then is equivalent to using the joint model anyway!

Acknowledgments

This research was supported by NIH National Institute of General Medical Science Grant GMS50011. This paper began as the UCLA Biostatistics PhD dissertation of the first author under supervision of the third author and continued as the UCLA Biostatistics MS report of the second author under the supervision of the third author. We thank Jeremy Taylor for useful comments.

References

Dixon, W. J. (1992). *BMDP Statistical Software Manual.* Los Angeles: University of California Press.

Fanurik, D. K. Zeltzer, M. C. Roberts, & R. L. Blount. (1993). The relationship between children's coping styles and psychological interventions for cold pressor pain. *Pain* **53**, 213–222.

Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Jennrich, R. I. & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820.

Louis, T. A. (1988). General Methods for Analysing Repeated Measures. *Statistics in Medicine* **7**, 29–45.

Reinsel, G. (1984). Estimation and prediction in a multivariate random-effects generalized linear model. *Journal of the American Statistical Association* **79**, 406–414.

SAS Institute Inc. (1996). *SAS/STAT Software, Changes and Enhancements.* Cary, NC: SAS Institute.

Shah, A., Laird, N. & Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92**, 775–779.

Sy, J. P., Taylor, J. M. G. & Cumberland, W. G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53**, 542–55.

Verbyla, A. P. & Cullis, B. R. (1990). Modelling in repeated measures experiments. *Applied Statistics* **39**, 341–56.

Weiss, R. E., Wang, Y. & Ibrahim, J. G. (1997). Predictive model selection for repeated measures random effects models using Bayes factors. *Biometrics* **53**, 159–69.

Parameter name	Symbol for log tolerance	Rating
intercept (attender)	α_{11}	α_{21}
baseline distractor - attender	α_{12}	α_{22}
AA	α_{13}	α_{23}
AD	α_{14}	α_{24}
AN	α_{15}	α_{25}
DA	α_{16}	α_{26}
DD	α_{17}	α_{27}
DN	α_{18}	α_{28}

Table 1: Definitions of regression parameters in the joint model. Labels AA, AD, etc. are two letter acronyms for coping style (A/D) treatment (A/D/N) combination.

Model	#	log like	AIC	test vs full chisq, df, p	test vs indep chisq, df, p	Selected p-values α_{12}	α_{17}	α_{24}
Prod Normal = Var	9	-648.5	-657.5	34.1, 27, .16	5.8, 1, .016	.01	.002	.03
Prod Normal $\neq$ Var	15	-646.4	-661.4	29.9, 21, .09	5.9, 1, .015	.01	.003	.04
Random Intercept	6	-655.2	-661.2	47.4, 30, .02	5.9, 2, .052	.02	.005	.25
MVAR(1), Diag A	5	-661.7	-666.7	60.4, 31, .001	11.2, 1, .001	.02	.01	.06
MVAR(1), Full A	7	-660.9	-667.9	58.9, 29, .001	12.7, 3, .005	.02	.01	.06
Unstructured	36	-631.5	-667.5	—	17.3, 16, .37	.01	.0006	.07

Table 2: Model: Model names for joint modeling approach; #: number of parameters in covariance structure; log likelihood; chisquare statistic, df, and p-value for testing the given model against the unstructured model; chisquare statistic, df, and p-value for testing the given model against the corresponding model of independence of Y_i and W_i; p-values for some significant α components. Except for the intercepts which we don't test, and α_{26} which had a p-value of under .0001, these are all variables that were significant in any model.

	M	Σ	#	Log like	AIC	test vs C5 chisq, df, p	test of $M = 0$ chisq, df, p
C1	Scalar	= Diag	8	-175.2	-183.2	17.84, 18, .47	4.6, 1, .03
C2	Diag	Unstruct	14	-169.6	-183.6	6.63, 12, .88	11.3, 4, .02
C3	CSYM	CSYM	4	-183.8	-187.8	35.18, 22, .08	6.0, 2, .05
C4	Lower	Unstruct	20	-168.3	-188.3	4.00, 6, .68	13.9, 10, .18
C5	Unstruct	Unstruct	26	-166.3	-192.3	—	17.9, 16, .33

Table 3: Results from fitting of conditional models (2). First column numbers the models, second column describes M where scalar means a constant times the identity; diag is a diagonal matrix; CSYM is compound symmetry $a_1 I + b_1 11'$ not necessarily with $a_1, b_1 > 0$; lower is lower triangular; and unstruct is unstructured. The next column, Σ, indicates the covariance structure for Y_i; = Diag is constant variance, with unstructured correlation matrix, CSYM is compound symmetry, restricted to being a covariance matrix, and unstruct means an unstructured covariance matrix. The # column is the number of parameters used in M and Σ, ignoring the 8 parameters in α. AIC is the log likelihood penalized by the number of parameters in M and Σ. The next column gives chisquare, df, and p-value for testing the given model against the unstructured M and Σ model C5. The last column gives the same information for the test of $H_0 : M = 0$.