UCSF
UC San Francisco Previously Published Works

Title
Joint modelling of mixed outcome types using latent variables.

Permalink
https://escholarship.org/uc/item/21g6d5jb

Journal
Statistical methods in medical research, 17(1)

ISSN
0962-2802

Author
McCulloch, Charles

Publication Date
2008-02-01

DOI
10.1177/0962280207081240

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Statistical Methods in Medical Research 2008; **17**: 53–73

Joint modelling of mixed outcome types using latent variables

Charles McCulloch Division of Biostatistics, Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

After a brief review of the use of latent variables to accommodate the correlation among multiple outcomes of mixed types, through theoretical and numerical calculation, the consequences of such a construction are quantified. The effects of including latent variables on marginal inference in these models are contrasted with the situation for jointly normal outcomes. A simulation study illustrates the efficiency and reduction in bias gains possible in using joint models, and analysis of an example from the field of osteoarthritis illustrates potential practical differences.

1 Introduction

Situations in which multiple outcomes are collected are common, but joint multivariate distributions that are sufficiently flexible to accommodate multiple outcomes and multiple predictors are rare. The multivariate normal distribution is by far the most commonly studied, even though many outcomes cannot successfully be approximated by a multivariate normal distribution.

Consider the situation where all the data from all the outcomes are stacked into a single data vector $\mathbf{Y}$, of order $\mathrm{N} \times 1$, and it is reasonable to assume it follows a multivariate normal distribution. A linear model can conceptually be built for the mean, leaving the variance–covariance matrix completely unspecified

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{V}). \tag{1}$$

However, with $\boldsymbol{Y}$ of size $\mathrm{N} \times 1$ there are $\mathrm{N}(\mathrm{N}-1)/2$ unique elements in $\boldsymbol{V}$. So there would be insufficient data to estimate all the elements of $\boldsymbol{V}$ empirically and some structure is needed. A common simplification is to model the variance–covariance structure using latent random variables. These are often called random effects[1] and those terms will be used interchangeably. That is, the distribution is first specified conditional on random latent variables, $\boldsymbol{b}$, which then induce a variance–covariance structure. A commonly used special case assumes that the conditional error term is independent and

Address for Correspondence: Charles McCulloch, Division of Biostatistics, Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA. E-mail: chuck@biostat.ucsf.edu

© 2008 SAGE Publications
Los Angeles, London, New Delhi and Singapore

10.1177/0962280207081240

homoscedastic

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{b} &\sim \mathcal{N}(\boldsymbol{X\beta} + \boldsymbol{Zb}, \boldsymbol{I}\sigma_e^2) \\ \boldsymbol{b} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}) \end{aligned} \tag{2}$$

where $\boldsymbol{Z}$ is a known model matrix for the random effects (playing the same role as $\boldsymbol{X}$ does for $\boldsymbol{\beta}$), $\boldsymbol{D}$ is the variance–covariance matrix of $\boldsymbol{b}$, which is usually of a much smaller dimension than $\boldsymbol{V}$ and the vertical bar denotes a conditional distribution. From this it is straightforward to derive the marginal distribution of Y

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{V}), \tag{3}$$

where $\boldsymbol{V} = \boldsymbol{ZDZ'} + \boldsymbol{I}\sigma_e^2$.

Several features of the model (2) and the progression from Equations (2) to (3) are worth noting

- The incorporation of random effects allows a flexible, but lower-dimensional specification of $\boldsymbol{V}$, the variance–covariance matrix of $\boldsymbol{Y}$.
- Inclusion of the random effects modifies only the variance–covariance matrix of $\boldsymbol{Y}$, not the marginal mean.
- Both the conditional and marginal distributions are multivariate normal.

Unfortunately, the multivariate normal linear model of the form Equation (1) does not generalize easily to other response types. This has led to the specification of models that naturally extend Equation (2) instead. For example, with binary outcome variables, logistic regression is commonly used and a natural generalization of Equation (2) is

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{b} &\sim \text{Bernoulli}(\boldsymbol{p}) \\ \ln(\boldsymbol{p}/[1-\boldsymbol{p}]) &= \boldsymbol{X\beta} + \boldsymbol{Zb} \\ \boldsymbol{b} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}), \end{aligned} \tag{4}$$

where the logarithm and division are performed elementwise on $\boldsymbol{p}$.

While this succeeds in the primary goal of building a correlated, multivariate distribution for the binary variables, it also leads to a more complicated situation since the incorporation of $\boldsymbol{b}$ also affects the marginal mean.[2,3] With binary variates, the only possible marginal distribution is Bernoulli, so the conditional and marginal distributions must be the same. However, this is not the case in general.

The idea of introducing random effects to incorporate correlation, as in Equation (4), is an old one, even for nonnormally distributed outcomes.[4] However, much less work has been done on the situation where the outcomes are of mixed types and measured on the same 'subjects' and there has been little quantification of the degree of association accommodated and the ramifications on the marginal distribution.

Much of the previous work has focussed on joint modelling of time-varying covariates and a time-to-event outcome in a survival analysis. The use of time-varying covariates

in a Cox model requires, ostensibly, the values of the covariates at all failure times. However, this is often not the case in practice. One solution has been to specify models with shared or correlated random effects for the covariates and time-to-event outcome so that they can be analysed together in a joint model. A number of authors have considered variations on such models.[5–10] The focus in such models is often on the time-to-event process rather than in joint modelling the multivariate outcomes, which is considered here.

Previous work that *has* been concerned with modelling joint outcomes has considered a variety of outcome types and fitting methods. Arminger and Küsters[4] specified a number of outcomes types arising from underlying continuous outcomes via thresholding (e.g., an ordinal categorical outcome arising from dividing a continuous outcome into ordered categories) and outlined approaches for maximum likelihood, without giving detailed estimation methods. Catalano and co-workers,[11–14] motivated by risk assessment problems, have considered a number of models, mostly based on shared and correlated random effects and have fit their models using both likelihood and generalized estimating equation approaches. Other modelling strategies include marginal modelling[15] fit via maximum likelihood, and Bayesian methods.[16,17] Models have been posited to deal with multivariate outcomes with continuous and binary[15] or continuous and ordinal[18] or mixed discrete outcomes.[17] Additional forms of correlation have also been accommodated, for example, multiple outcomes with clustered data,[12] as in data from animal litters in toxicology experiments, and time series data,[19] as in economic data.

Two motivating examples and some of the rationale for considering joint models are described first. In Section 4, the effects of adding shared random effects to a probit/normal example and a Poisson/normal example are quantified. Sections 5–7 briefly consider correlated latent variables, discrete versus continuous latent variables and missing data. Section 8 presents a simulation study, Section 9 returns to the analysis of one of the motivating examples and Section 10 concludes with a summary.

2 Motivating examples

To motivate the later developments, two examples are briefly described in medical research.

2.1 Example: Medical services utilization

A small percentage of patients treated by a hospital system use most of the resources – often in ways that can be prevented. For example, persons without insurance may use the emergency room for non-emergency care. A randomized trial of 190 patients was conducted[20,21] to test if a managed care intervention was able to improve access to healthcare when compared with standard care. Measurements were taken at baseline, 6 months, 12 months and 18 months after randomization. Outcomes included cost of care, number of emergency room visits and death. Predictors included treatment group (managed care or not), gender, the Beck Depression Inventory[22] and whether the person was homeless. A primary focus was on the treatment effect, while adjusting for the effects

of the other predictors. A secondary goal was to assess the impact of all the predictors on the outcomes.

2.2 Example: Osteoarthritis Initiative

The Osteoarthritis Initiative (OAI) is a multicenter cohort study, conducted in part to understand risk factors for progression of osteoarthritis (OA) of the knee. OA causes problems ranging from stiffness and mild pain to severe joint pain and even disability. Approximately 4000 men and women aged $\geq$45 at high risk for developing knee OA will be followed yearly for four years. Two of the outcomes collected are the Western Ontario and McMaster Universities (WOMAC) disability score,[23] which is usually modelled as a continuous variable, and number of days of work missed out of the last three months due to knee pain, which is a count variable. Data from the OAI are available for download from the website: http://www.oai.ucsf.edu/datarelease/.

3 Advantages of joint modelling

When faced with multiple outcomes in a data analysis, a reasonable first step is to analyse each outcome separately. However, there are a number of reasons that the joint modelling of outcomes is advantageous.

3.1 Accommodate multiple outcomes

In studies that collect multiple outcomes, scientific interest may focus naturally on the simultaneous occurrence of one or more outcomes. For example, Catalano *et al.*[11] develop statistical models for quantitative risk assessment of the influence of possibly toxic agents in animal experiments. Administration of toxic agents can cause foetal death, malformation at birth, lack of growth after birth or neurological deficits in any of a number of domains. It would be artificial to analyse these separately – for example, an agent that caused 40% foetal death and no neurological deficits would appear better than an agent causing 5% foetal death and 25% neurological deficits when analysing the separate outcome of neurological deficit (either among all animals or among survivors). This type of conclusion is not useful for assessment of risk, whereas a joint analysis allows the assessment of the overall impact as well as the separate and joint effects of the agent on all the outcomes. In the example of Section 2.1 interest naturally focuses on the simultaneous impact of the intervention on all the outcomes.

3.2 Avoid multiple testing

By forming joint models, it is straightforward to calculate an overall test of the effect of a predictor without having to resort to ad hoc methods such as Bonferroni adjustment. In the risk assessment arena, interest might focus on the effect of a toxic agent on any or all of the outcomes. In a likelihood-based analysis, this can easily be achieved by forming a likelihood ratio test.

3.3 Model correlation between outcomes

In other settings, the correlation between the outcomes is of interest and thus forces a joint analysis to model that correlation. For example, in the OAI, the WOMAC pain score is calculated from an easily collected questionnaire and it is of interest to see how well it correlates with much more expensive and involved magnetic resonance imaging evaluation of OA.[24]

3.4 Efficiency gain

Separate analysis of outcomes can be inefficient. Section 8 shows significant efficiency gains, which can be achieved in certain situations.

3.5 Handle missing at random data

When the data on one of the outcomes are more complete than another then joint analysis can accommodate data that are missing at random[25] instead of the stronger assumption of missing completely at random (MCAR). For example, in the OAI, subjects are not officially enrolled in the study unless their baseline data are relatively complete. However, data collected at later time points (e.g., total knee replacement at year 3) will only be available for subjects who have not dropped out of the study. And the likelihood that they drop out may be related to the values of baseline variables (e.g., WOMAC pain score). This idea is developed more fully in Section 7.

In a different vein, models can be created where one of the outcomes is itself whether the data are missing or not (either as binary or as a time between measurements). Wu and Carroll[5] adopt this approach with a continuous longitudinal outcome of interest using a random effect shared between the longitudinal outcome and the dropout process.

4 Consequences of shared latent variables or random effects

While there are often good reasons to model the outcomes jointly, the specification of joint models is not innocuous. In this section, the use of shared random effects to build joint outcome models is illustrated, akin to Equation (4). Two joint models are developed: a probit/normal model and a Poisson/normal model and they are used to elaborate on consequences of the specification of models.

The basic idea is to use a random effect to build in a correlation between the two outcomes. To fix ideas, consider the OAI example with two outcomes: Y_{1i} being the binary outcome of total knee replacement at three years and Y_{2i} being the outcome of WOMAC disability score at baseline, both for subject i. If we were modelling the outcomes separately we might build a probit model (the reason for using probit rather than logit will become clear later) for Y_1 and a linear model for Y_2 and we will consider just a single binary predictor $x =$ use of glucosamine/chondroitin supplement or not. So, the separate models would be

$$Y_{1i} \sim \text{indep. Bernoulli}(\nu_i^B)$$
$$\Phi^{-1}(\nu_i^B) = \gamma_0^B + \gamma_1^B x_i$$

$$\begin{aligned} Y_{2i} &\sim \text{indep. } \mathcal{N}(\nu_i^N, \tau^2) \\ \nu_i^N &= \gamma_0^N + \gamma_1^N x_i, \end{aligned} \tag{5}$$

and the superscripts indicate that the parameters are specific to the binary (B) or normal (N) models.

While Equation (5) would suffice for separate analyses, it does not accommodate a correlation between Y_{1i} and Y_{2i}. A simple device to do so is to introduce a random effect that will be shared by both responses for any particular subject. Equation (5) is modified accordingly by modelling the distributions conditional on the random effect, b_i

$$\begin{aligned} Y_{1i}|b_i &\sim \text{indep. Bernoulli}(\mu_i^B) \\ \Phi^{-1}(\mu_i^B) &= \beta_0^B + \beta_1^B x_i + b_i \\ Y_{2i}|b_i &\sim \text{indep. } \mathcal{N}(\mu_i^N, \sigma^2) \\ \mu_i^N &= \beta_0^N + \beta_1^N x_i + \lambda b_i, \\ b_i &\sim \text{ i.i.d. } G, \end{aligned} \tag{6}$$

where, as of yet, G will remain unspecified and we make the assumption that, conditional on b_i, Y_{1i} and Y_{2i} are *conditionally independent*. The λ multiplying b_i in the equation for Y_{2i} is to account for the fact that the linear predictors for Y_1 and Y_2 are measured on different scales and it is unreasonable to assume they have the same variance. The consequences of such a specification are now explored.

4.1 Conditional independence

With only two outcomes, the assumptions behind the conditional independence are rather mild (since there is only a single association to be modeled). When conditional independence holds two identities are particularly useful, the iterated conditional expectation identity and a similar identity for covariances

$$E[Y_{ji}] = E[E[Y_{ji}|b_i]] \tag{7}$$

$$\text{cov}(Y_{ji}, Y_{lk}) = E[\text{cov}(Y_{ji}, Y_{lk}|b_i)] + \text{cov}(E[Y_{ji}|b_i], E[Y_{lk}|b_i]), \tag{8}$$

with Equation (8) simplifying when conditional independence holds because, for example, $\text{cov}(Y_{1i}, Y_{2i}|b_i) = 0$. These identities are used to check the consequences of adding the random effect in Equation (6).

4.1.1 *Marginal means*

The marginal mean for Y_2 from Equation (6), averaging over the distribution of the random effects, is simply calculated using Equation (7)

$$\begin{aligned} E[Y_{2i}] &= E[\beta_0^N + \beta_1^N x_i + \lambda b_i] \\ &= \beta_0^N + \beta_1^N x_i + \lambda E[b_i]. \end{aligned} \tag{9}$$

Since the mean of b_i can be absorbed into β_0^N it is often convenient to assume that $E[b_i]=0$, which corresponds to modelling the individual specific random effects as deviations from the overall intercept, β_0^N. This simplifies the marginal mean to $\beta_0^N + \beta_1^N x_i$, the same as the mean in the separate model, Equation (5), that is, $\gamma_0^N = \beta_0^N$ and $\gamma_1^N = \beta_1^N$. Even if the mean of b_i is not assumed to be 0, the regression coefficients of x, γ_1^N and β_1^N, from Equations (5) and (6) will be equal.

For Y_1, the results are a bit more complicated and the simplifying assumption that $b_i \sim \mathcal{N}(0, \sigma_b^2)$ is made. For this calculation, note that $\Phi(X) = \Pr\{Z < X|X\}$, where $Z \sim \mathcal{N}(0,1)$, independent of X and Equation (7) is again used

$$\begin{aligned}
E[Y_{1i}] &= E\left[\Phi(\beta_0^B + \beta_1^B x_i + b_i)\right] \\
&= E\left[\Pr\{Z < \beta_0^B + \beta_1^B x_i + b_i | b_i\}\right] \\
&= \Pr\left\{Z < \beta_0^B + \beta_1^B x_i + b_i\right\} \\
&= \Pr\left\{Z - b_i < \beta_0^B + \beta_1^B x_i\right\} \\
&= \Pr\left\{\frac{Z - b_i}{\sqrt{1+\sigma_b^2}} < \frac{\beta_0^B + \beta_1^B x_i}{\sqrt{1+\sigma_b^2}}\right\} \\
&= \Phi\left(\frac{\beta_0^B + \beta_1^B x_i}{\sqrt{1+\sigma_b^2}}\right) \\
&= \Phi\left(\gamma_0^B + \gamma_1^B x_i\right), \qquad (10)
\end{aligned}$$

where the third identity holds because the expected value of the conditional probability is the unconditional probability.

There are several things to note associated with Equation (10). First, the marginal regression model is available in 'closed form' and is again a probit model (this is the reason for using the probit rather than the logit). Secondly, the last identity in Equation (10) holds because the separate outcomes model is a model for the marginal mean of Y_1. Thirdly, the marginal regression coefficients are not the same as the conditional coefficients and, in particular, are smaller by a factor of $\sqrt{1+\sigma_b^2}$,

$$\gamma_p^B = \beta_p^B / \sqrt{1+\sigma_b^2}. \qquad (11)$$

An important consequence of this is that the regression parameters for the binary outcome will not directly be comparable between the separate and joint outcomes models. However, for a binary probit model the adjustment is straightforward and for logit models a good approximation is available.[26]

4.1.2 Marginal variances and covariance

Responses Y_{2i} for different subjects are assumed to be independent. The variance can be calculated using Equation (8)

$$\begin{aligned}\text{cov}(Y_{2i}, Y_{2i}) &= E[\text{cov}(Y_{2i}, Y_{2i}|b_i)] + \text{cov}(E[Y_{2i}|b_i], E[Y_{2i}|b_i]) \\ &= E[\text{var}(Y_{2i}|b_i)] + \text{cov}(E[Y_{2i}|b_i], E[Y_{2i}|b_i]) \\ &= \sigma^2 + \text{cov}(\beta_0^N + \beta_1^N x_i + \lambda b_i, \beta_0^N + \beta_1^N x_i + \lambda b_i) \\ &= \sigma^2 + \lambda^2 \sigma_b^2, \end{aligned} \tag{12}$$

so that τ^2 of Equation (5) is given by $\sigma^2 + \lambda^2\sigma_b^2$. The interpretation of this result is that part of the marginal variance in Y_{2i} is being attributed to variation in the shared random effect.

The Y_{1i} are similarly assumed to be independent and, though long-winded as described in the next section, the same strategy can be adopted for calculating the variance, recalling the fact, Equation (10), that $E[Y_{1i}] = E[\Phi(\mu_i^B)] = \Phi(\nu_i^B)$

$$\begin{aligned}\text{cov}(Y_{1i}, Y_{1i}) &= E[\text{cov}(Y_{1i}, Y_{1i}|b_i)] + \text{cov}(E[Y_{1i}|b_i], E[Y_{1i}|b_i]) \\ &= E[\text{var}(Y_{1i}|b_i)] + \text{cov}(E[Y_{1i}|b_i], E[Y_{1i}|b_i]) \\ &= E\left[\Phi\left(\mu_i^B\right)\left\{1 - \Phi\left(\mu_i^B\right)\right\}\right] + \text{var}\left(\Phi\left(\mu_i^B\right)\right) \\ &= E\left[\Phi\left(\mu_i^B\right)\right] - E\left[\Phi^2\left(\mu_i^B\right)\right] + E\left[\Phi^2\left(\mu_i^B\right)\right] - E\left[\Phi\left(\mu_i^B\right)\right]^2 \\ &= \Phi\left(\nu_i^B\right) - \Phi^2\left(\nu_i^B\right) \\ &= \Phi\left(\nu_i^B\right)\left[1 - \Phi\left(\nu_i^B\right)\right]. \end{aligned} \tag{13}$$

Using the same identity, the covariance is calculated as

$$\begin{aligned}\text{cov}(Y_{1i}, Y_{2i}) &= E[\text{cov}(Y_{1i}, Y_{2i}|b_i)] + \text{cov}(E[Y_{1i}|b_i], E[Y_{2i}|b_i]) \\ &= 0 + \text{cov}(\Phi(\beta_0^B + \beta_1^B x_i + b_i), \beta_0^N + \beta_1^N x_i + \lambda b_i) \\ &= \text{cov}(\Phi(\beta_0^B + \beta_1^B x_i + b_i), \lambda b_i). \end{aligned} \tag{14}$$

Letting $\theta_i^B = \beta_0^B + \beta_1^B x_i$, this can be rewritten as

$$\begin{aligned}\text{cov}(Y_{1i}, Y_{2i}) &= \text{cov}(\Phi(\theta_i^B + b_i), \lambda b_i) \\ &= \lambda \int_{-\infty}^{\infty} \sigma_b z \Phi(\theta_i^B + \sigma_b z)\phi(z)\text{d}z, \end{aligned} \tag{15}$$

where $\phi(z)$ is the standard normal p.d.f. Although this cannot be evaluated in closed form, it is not too hard to evaluate numerically.

Using Equation (15) and the variance calculations, the correlation is given by

$$\text{corr}(Y_{1i}, Y_{2i}) = \sqrt{\frac{\lambda^2\sigma_b^2}{\lambda^2\sigma_b^2 + \sigma^2}} \frac{\int_{-\infty}^{\infty} z\Phi(\theta_i^B + \sigma_b z)\phi(z)\mathrm{d}z}{\sqrt{\Phi\left(\theta_i^B/\sqrt{1+\sigma_b^2}\right)\left[1 - \Phi\left(\theta_i^B/\sqrt{1+\sigma_b^2}\right)\right]}}. \quad (16)$$

The leading term under the square root sign in Equation (16) would be the familiar intraclass correlation coefficient, $\rho = \lambda^2\sigma_b^2/(\lambda^2\sigma_b^2 + \sigma^2)$, if both Y_1 and Y_2 were bivariate normally distributed (recall that the variance of the random effect in the normal distribution linear predictor was $\lambda^2\sigma_b^2$).

Table 1 gives the correlation, Equation (16), for various values of θ^B and ρ, first for $\sigma^2 = 1$ and then for $\sigma^2 = 2$. Several features of Table 1 are noteworthy

- The value of the correlation closely tracks the value of the 'intraclass correlation,' ρ for a given value of θ^B.
- The values are sensitive to both ρ and θ^B.
- The values are not overly sensitive to the value of σ^2 for a fixed ρ.
- The correlation does *not* increase toward 1 as ρ increases or, equivalently, as σ_b^2 increases for a fixed value of σ^2.

In fact, for a fixed value of σ^2 it is straightforward to take the limit as σ_b tends to infinity

$$\lim_{\sigma_b \to \infty} \text{corr}(Y_{1i}, Y_{2i}) = \sqrt{1}\frac{\int_0^{\infty} z\phi(z)\mathrm{d}z}{\sqrt{\Phi(0)[1 - \Phi(0)]}} = \sqrt{\frac{2}{\pi}} \approx 0.798. \quad (17)$$

4.1.3 Marginal distributions

The marginal distributions are simple to work out for Equation (6). Since the only possible distribution for binary random variables is Bernoulli, the marginal distribution of Y_{1i} is Bernoulli with success probability given by Equation (10). This provides a

Table 1 Correlations for a probit/normal model

θ^B	ρ					
	0	0.1	0.3	0.5	0.7	0.9
$\sigma^2 = 1$						
0	0.00	0.08	0.24	0.40	0.56	0.72
1	0.00	0.07	0.21	0.36	0.53	0.71
2	0.00	0.04	0.14	0.27	0.45	0.68
3	0.00	0.01	0.07	0.16	0.33	0.61
$\sigma^2 = 2$						
0	0.00	0.11	0.30	0.46	0.61	0.74
1	0.00	0.09	0.27	0.43	0.59	0.73
2	0.00	0.06	0.20	0.36	0.53	0.71
3	0.00	0.02	0.11	0.26	0.45	0.67

simpler method to calculate the variance, as opposed to Equation (13), as the product of the marginal mean, Equation (10), and one minus the marginal mean.

The marginal distribution of Y_{2i} is normal with the mean and variance given by Equations (9) and (12), respectively. Having the distribution closed under addition of random effects is somewhat unusual as described in Section 4.2.

4.1.4 Variations on the probit/normal model

Various modifications and elaborations of the probit/normal model are possible. Many of these center on the idea of a threshold model: an underlying, normally distributed variable $W_i \sim \mathrm{N}(\beta_0^W + \beta_1^W x_i, 1)$ is defined and recording whether it exceeds a threshold of 0 gives rise to a probit model.

$$\begin{aligned}\Pr(W_{1i} > 0) &= \Pr\left(W_{1i} - \left[\beta_0^W + \beta_1^W x_i\right] > -\left[\beta_0^W + \beta_1^W x_i\right]\right) \\ &= 1 - \Phi\left(-\left[\beta_0^W + \beta_1^W x_i\right]\right) \\ &= \Phi\left(\beta_0^W + \beta_1^W x_i\right).\end{aligned} \tag{18}$$

That is, a probit model arises as long as the event $W_{1i} > 0$ is identified with $Y_{1i} = 1$. This device extends easily to ordinal outcomes by introducing several threshold values. This underlying construction has been used by a number of authors to build correlated data models.[4,18]

The calculation Equation (16) is reminiscent of the tetrachoric correlation, which describes the correlation between two binary outcomes under a threshold model for each. Qu *et al*.[27] describe the relationships between latent variable models and the tetrachoric correlation.

4.2 A Poisson/normal example

A slightly different model is used to illustrate both the lack of closure of the distribution and another feature of shared random effects models. In place of the binary variate, a joint model with conditional (on the random effects) Poisson and normal distributions is considered, using the canonical links for those two distributions (log and identity, respectively)

$$\begin{aligned} Y_{1i}|b_i &\sim \text{indep. Poisson}(\mu_i^C) \\ \log(\mu_i^C) &= \beta_0^C + \beta_1^C x_i + b_i \\ Y_{2i}|b_i &\sim \text{indep. } \mathcal{N}(\mu_i^N, \sigma^2) \\ \mu_i^N &= \beta_0^N + \beta_1^N x_i + \lambda b_i \\ b_i &\sim \text{ i.i.d. } \mathcal{N}(0, \sigma_b^2). \end{aligned} \tag{19}$$

The marginal distribution of Y_{2i} is the same as for model (6), so attention is focussed on the distribution of Y_{1i}.

4.2.1 Marginal means

The marginal mean for Y_1 from Equation (19) is again calculated from Equation (7)

$$\begin{aligned} E[Y_{1i}] &= E[e^{\beta_0^C+\beta_1^C x_i+b_i}] \\ &= e^{\beta_0^C+\beta_1^C x_i} E[e^{b_i}] \\ &= e^{\beta_0^C+\beta_1^C x_i} M_{b_i}(1), \end{aligned} \tag{20}$$

where $M_W(t)$ is the moment generating function of the random variable W evaluated at t. It is assumed that $b_i \sim \mathcal{N}(0, \sigma_b^2)$ so that $M_{b_i}(1) = e^{\sigma_b^2/2}$ and therefore

$$E[Y_{1i}] = \exp\left\{\beta_0^C + \beta_1^C x_i + \frac{\sigma_b^2}{2}\right\}. \tag{21}$$

Of note is that the log of the marginal mean is $\beta_0^C + \beta_1^C x_i + \sigma_b^2/2$, which is the same as the log of the conditional mean, except offset by $\sigma_b^2/2$. In particular, the regression coefficient for x_i, β_1^N is the same in both the marginal and conditional models.

4.2.2 Marginal variance and covariance

The Y_{1i} are assumed to be independent with the variance calculated by the usual formula

$$\begin{aligned} \text{cov}\,(Y_{1i}, Y_{1i}) &= E\left[\text{var}\,(Y_{1i}|b_i)\right] + \text{cov}\left(E\left[Y_{1i}|b_i\right], E\left[Y_{1i}|b_i\right]\right) \\ &= E\left[E\left[Y_{1i}|b_i\right]\right] + \text{var}\left(e^{\beta_0^C+\beta_1^C x_i} e^{b_i}\right) \\ &= E\left[Y_{1i}\right] + e^{2\beta_0^C+2\beta_1^C x_i}\text{var}\left(e^{b_i}\right) \\ &= E\left[Y_{1i}\right] + e^{2\beta_0^C+2\beta_1^C x_i}\left(M_{b_i}(2) - M_{b_i}(1)^2\right) \\ &= E\left[Y_{1i}\right] + e^{2\beta_0^C+2\beta_1^C x_i}\left(e^{2\sigma_b^2} - e^{\sigma_b^2}\right) \\ &= E\left[Y_{1i}\right] + E\left[Y_{1i}\right]^2\left(e^{\sigma_b^2} - 1\right). \end{aligned} \tag{22}$$

The covariance is calculated as before

$$\begin{aligned} \text{cov}(Y_{1i}, Y_{2i}) &= E\left[\text{cov}(Y_{1i}, Y_{2i}|b_i)\right] + \text{cov}\left(E[Y_{1i}|b_i], E[Y_{2i}|b_i]\right) \\ &= 0 + \text{cov}(e^{\beta_0^C+\beta_1^C x_i} e^{b_i}, \beta_0^N + \beta_1^N x_i + \lambda b_i) \\ &= \lambda e^{\beta_0^C+\beta_1^C x_i}\text{cov}(e^{b_i}, b_i). \end{aligned} \tag{23}$$

The last term in this calculation can be evaluated as

$$\begin{aligned}\text{cov}(e^{b_i}, b_i) &= E[b_i e^{b_i}] \\ &= E[\sigma_b Z e^{\sigma_b Z}], \quad \text{with } Z \sim \mathcal{N}(0,1) \\ &= \sigma_b^2 e^{\sigma_b^2/2}. \end{aligned} \tag{24}$$

Combining Equations (23) and (24) gives

$$\begin{aligned}\text{cov}(Y_{1i}, Y_{2i}) &= \lambda e^{\beta_0^C + \beta_1^C x_i} \sigma_b^2 e^{\sigma_b^2/2} \\ &= \lambda E[Y_{1i}] \sigma_b^2. \end{aligned} \tag{25}$$

Therefore, the correlation between Y_{1i} and Y_{2i} is

$$\text{corr}(Y_{1i}, Y_{2i}) = \frac{E[Y_{1i}]\sigma_b^2}{\sqrt{(\sigma^2/\lambda^2 + \sigma_b^2)\left(E[Y_{1i}] + E[Y_{1i}]^2(e^{\sigma_b^2} - 1)\right)}}. \tag{26}$$

Table 2 gives the correlation for various values of $\theta_i^C = \beta_0^C + \beta_1^C x_i$ and $\rho = \sigma_b^2/(\sigma_b^2 + \sigma^2)$ with $\sigma^2 = \lambda^2$. Table 2 reflects the following

- The value of the correlation does *not* increase monotonically as σ_b increases.
- In fact, as σ_b increases the correlation has a limiting value of zero.
- The correlation tends to a limit as θ^C increases, for a fixed value of ρ.

For a fixed value of the mean, the limit of Equation (26) is governed by the limit of $\sigma_b^4/(e^{\sigma_b^2} - 1)$, which tends to zero. For a fixed value of σ_b^2, the limit of the correlation, as θ^C increases, is easily seen to be $\sigma_b^2/\sqrt{(\sigma^2/\lambda^2 + \sigma_b^2)(e^{\sigma_b^2} - 1)}$.

The intuitive explanation for why the correlation tends to zero as σ_b^2 increases is that the variability of the Poisson distribution increases faster than the covariance, driving the correlation to zero. This arises because, for large σ_b, the distribution is highly overdispersed.

Table 2 Correlations for a Poisson/normal model with $\sigma^2 = \lambda^2$

θ^C	ρ					
	0	0.1	0.3	0.5	0.7	0.9
0	0.00	0.10	0.31	0.46	0.41	0.03
1	0.00	0.15	0.39	0.51	0.42	0.03
2	0.00	0.21	0.45	0.53	0.42	0.03
3	0.00	0.26	0.47	0.53	0.42	0.03

4.2.3 Marginal distributions

In fact, the overdispersion of the marginal distribution of Y_{2i} is easily calculated from Equation (23) as

$$\begin{aligned}\text{overdispersion} &= \frac{\text{var}(Y_{1i})}{E[Y_{1i}]} \\ &= 1 + E[Y_{1i}](e^{\sigma_b^2} - 1). \end{aligned} \tag{27}$$

Since the overdispersion is greater than 1 whenever the random effects have variance greater than 0, the marginal distribution of Y_{1i} is always overdispersed compared to a Poisson distribution. So, in this case the introduction of a shared random effect changes the marginal distribution. Furthermore, the amount of overdispersion (a characteristic of the marginal distribution of Y_{1i}) is intimately associated with the correlation between Y_{1i} and Y_{2i}.

5 Correlated random effects

Previous sections have demonstrated that the use of shared random effects has the advantage of building a positive correlation between random variables with different distributions. However, the flexibility of the approach does not match that of the multivariate normal distribution, which has the nice feature that the inclusion of random effects modifies only the variance-covariance structure. With distributions of mixed types, especially the binomial and Poisson distributions, in which the variance is a function of the mean, the introduction of shared random effects can have multifactorial consequences, some of which may not be advantageous, for example, the connection between the overdispersion and the correlation observed in the previous section.

A fairly straightforward generalization of a shared random effect involves the introduction of correlated random effects. This idea is illustrated by considering a generalization of the Poisson/normal model, Equation (26), in which separate, but correlated random effects are allowed in the two models

$$\begin{aligned} Y_{1i}|b_i &\sim \text{indep. Poisson}(\mu_i^C) \\ \log(\mu_i^C) &= \beta_0^C + \beta_1^C x_i + b_i^C \\ Y_{2it}|b_i &\sim \text{indep. } \mathcal{N}(\mu_i^N, \sigma^2) \\ \mu_i^N &= \beta_0^N + \beta_1^N x_i + b_i^N \\ \boldsymbol{b}_i &= \begin{pmatrix} b_i^C \\ b_i^N \end{pmatrix} \sim \text{ i.i.d. } \mathcal{N}_2(0, \boldsymbol{\Sigma}_b), \end{aligned} \tag{28}$$

where repeated measurements over time on the continuous outcome, Y_{2it} $(t = 1, \ldots, n_i)$ are assumed in order to identify the subject to subject variation, $\text{var}(b_i^N)$.

Using the notation

$$\Sigma_b = \begin{pmatrix} \sigma_{bc}^2 & \sigma_{bc}\sigma_{bn}\rho_{cn} \\ \sigma_{bc}\sigma_{bn}\rho_{cn} & \sigma_{bn}^2 \end{pmatrix}, \tag{29}$$

the marginal means and variances are little changed

$$E[Y_{1i}] = \exp\left\{\beta_0^C + \beta_1^C x_i + \frac{\sigma_{bc}^2}{2}\right\} \tag{30}$$

$$\text{var}(Y_{1i}) = E[Y_{1i}] + E[Y_{1i}]^2(e^{\sigma_{bc}^2} - 1) \tag{31}$$

$$E[Y_{2it}] = \beta_0^N + \beta_1^N x_i \tag{32}$$

$$\text{var}(Y_{2it}) = \sigma^2 + \sigma_{bn}^2. \tag{33}$$

The covariance calculation is somewhat different

$$\begin{aligned} \text{cov}\,(Y_{1i}, Y_{2it}) &= E\left[\text{cov}\,(Y_{1i}, Y_{2it}|b_i)\right] + \text{cov}\left(E\left[Y_{1i}|b_i\right], E\left[Y_{2it}|b_i\right]\right) \\ &= 0 + \text{cov}\left(e^{\beta_0^C + \beta_1^C x_i} e^{b_i^C}, \beta_0^N + \beta_1^N x_i + b_i^N\right) \\ &= e^{\beta_0^C + \beta_1^C x_i} \text{cov}\left(e^{b_i^C}, b_i^N\right). \end{aligned} \tag{34}$$

The calculation of the final term is somewhat easier if the correlated random effects are rewritten in terms of three i.i.d. standard normal variates, Z_i

$$\begin{aligned} b_i^C &= \sigma_{bc}\left(Z_1\sqrt{1-|\rho_{cn}|} + Z_3\sqrt{|\rho_{cn}|}\right) \\ b_i^N &= \sigma_{bn}\left(Z_2\sqrt{1-|\rho_{cn}|} + Z_3\,\text{sgn}(\rho_{cn})\sqrt{|\rho_{cn}|}\right). \end{aligned}$$

Using this representation, the final term of the covariance is calculated as

$$\begin{aligned} \text{cov}(e^{b_i^C}, b_i^N) &= E[b_i^N e^{b_i^C}] \\ &= E\left[\sigma_{bn}\left(Z_2\sqrt{1-|\rho_{cn}|} + Z_3\,\text{sgn}(\rho_{cn})\sqrt{|\rho_{cn}|}\right) e^{\sigma_{bc}(Z_1\sqrt{1-|\rho_{cn}|}+Z_3\sqrt{|\rho_{cn}|})}\right], \\ &= \sigma_{bn}\,\text{sgn}(\rho_{cn})\sqrt{|\rho_{cn}|}E\left[Z_3 e^{\sigma_{bc}(Z_1\sqrt{1-|\rho_{cn}|}+Z_3\sqrt{|\rho_{cn}|})}\right], \\ &= \sigma_{bn}\,\text{sgn}(\rho_{cn})\sqrt{|\rho_{cn}|}E\left[Z_3 e^{\sigma_{bc}Z_3\sqrt{|\rho_{cn}|}}\right], \\ &= \sigma_{bn}\,\text{sgn}(\rho_{cn})\sqrt{|\rho_{cn}|}\sigma_{bc}\sqrt{|\rho_{cn}|}e^{\sigma_{bc}^2|\rho_{cn}|/2}, \\ &= \sigma_{bn}\sigma_{bc}\rho_{cn}e^{\sigma_{bc}^2|\rho_{cn}|/2}. \end{aligned} \tag{35}$$

Combining Equations (35) and (34) gives the covariance as

$$\text{cov}(Y_{1i}, Y_{2it}) = e^{\beta_0^C + \beta_1^C x_i} \sigma_{bn} \sigma_{bc} \rho_{cn} e^{\sigma_{bc}^2 |\rho_{cn}|/2}, \tag{36}$$

and a correlation of

$$\text{corr}(Y_{1i}, Y_{2it}) = \frac{e^{\beta_0^C + \beta_1^C x_i} \sigma_{bn} \sigma_{bc} \rho_{cn} e^{\sigma_{bc}^2 |\rho_{cn}|/2}}{\sqrt{(\sigma^2 + \sigma_{bn}^2)\left(E[Y_{1i}] + E[Y_{1i}]^2 (e^{\sigma_{bc}^2} - 1)\right)}}. \tag{37}$$

The correlation given by Equation (37) gives somewhat more flexibility in modelling. First and importantly, it allows negative correlations between the two outcomes. Secondly, it gives more latitude in disentangling the overdispersion, which is governed by σ_{bc}^2 and the covariance, which is additionally governed by ρ_{cn}.

5.1 Random effects versus correlated errors

Model (28) is constructed by adding random effects to the linear predictors to incorporate correlation. The linear model, with its assumed normal distribution, has an additional random 'error' term, with variance σ^2. An alternate parameterization is to add 'error' terms to both models and then assume a multivariate normal distribution. An advantage to this approach is that it opens up the full range of modelling of the covariance structure associated with the multivariate normal distribution. For example, this was used to advantage to incorporate time series correlation in a count data model.[28]

When the distribution is Poisson, this may have unintended consequences, since it forces overdispersion, similar to Equation (31). However, with models built from underlying latent normal variables, as discussed in Section 4.1.4, the approach is quite attractive.

6 Discrete versus continuous random effects

In the previous section it was assumed, in order to obtain closed form expressions, that the random effects were normally or multivariate normally distributed. However, many of the results hold irrespective of the specific distributional assumption for the random effects. For example, Equation (20) is completely general, as is the result that the conditional regression coefficient for x_i is the same as the marginal regression coefficient since

$$\log E[Y_{1i}] = \beta_0^C + \beta_1^C x_i + \log M_{b_i}(1). \tag{38}$$

In particular, a sometimes convenient assumption for the random effects distribution is a discrete one. This converts the integrals in, for example, Equation (24) to be summations, which can computationally be expedient.

Discrete random variables for the random effects can also be compatible with interpretations in terms of unobserved sub-populations. For example, consider the medical

resource utilization example from Section 2.1. If such an analysis was performed from administrative data, it might not have information as to whether the person was homeless or not. It would be easy to imagine that both death rates and number of emergency room visits were higher in the homeless group, which would generate correlated death and emergency room visit outcomes.

Discrete random variables can also be useful for capturing variation over and above that usually described using continuous latent variables. For example, in a longitudinal study like the medical services example, continuous (and usually normally distributed) random effects are often used to model the correlated nature of the longitudinal measurements over time within a person, but then an observed categorical variable like homeless status could generate more extensive heterogeneity. McCulloch and co-workers[10,29] explore such applications.

7 Missing data

As mentioned in Section 3, an advantage of the joint modelling approach is the ability to accommodate situations in which the data are missing in a way predictable by one of the outcomes. More specifically, suppose that primary interest focuses on Y_1, but that the dropout process depends on Y_2. For example, in the OAI, those with higher WOMAC pain scores at baseline (Y_2) may have a higher likelihood of dropping out and therefore not furnishing a value for Y_1, the occurrence of total knee replacement at year 3.

Because Y_1 and Y_2 are correlated, if a separate, marginal model (e.g., logistic regression) is used to analyze Y_1, then the data are not missing at random and inconsistent estimators can result. On the other hand, if the missing process depends on Y_2 and the data are analysed by maximum likelihood, then the argument of Laird[30] can easily be applied to joint models with shared or correlated random effects to show that consistent estimators will result.

8 A simulation study

A simulation study was performed to demonstrate the feasibility of fitting shared random effects models and to explore more quantitatively the benefits of such models. Data were simulated from the probit/normal model, Equation (6), with a binary predictor x_i, equal to 0 or 1, equal sample sizes for each group and a normal distribution for the random effects. The focus of the simulation was estimation of the regression coefficients, β_1^B and β_1^N. The overall sample size was N = 100 for each outcome. Four simulation scenarios were generated: no missing data, 15% of the data missing completely at random and 15% of the data missing and dependent (either mildly or severely) on the continuous outcome Y_2. Two thousand replications were performed and evaluated under each scenario. The simulation was implemented as a do-file in Stata Release 9.1 (Stata Corporation, College Station, TX, USA) and the shared random effect models were fit using GLLAMM (www.gllamm.org), which uses adaptive quadrature[31] to fit the models by maximum likelihood.

The values of the parameters were $\beta_0^B = \beta_0^N = 0, \beta_1^B = 1, \beta_1^N = 2, \sigma^2 = \sigma_b^2 = 1$. The two missing at random processes were simulated by dropping subjects with small values of Y_{2i} at higher rates than those with large values. In the first, more extreme, scenario subjects with below median values of Y_{2i} dropped out at a rate of 28.7% whereas those with above median values of Y_{2i} dropped out at a rate of 1.3%. The exact mechanism was to add a uniform random variable to a multiple of Y_{2i} and delete those with lower values of the sum. In the less severe scenario, the dropout rates were 21.8 and 8.2%, respectively. The multipliers of Y_{2i} used to achieve these rates were, respectively, 0.15 and 0.05.

For each data set, the joint probit/normal model was fit as well as separate probit and linear regression models. Because the marginal and conditional regression coefficients are not equal, the conditional coefficient was adjusted to the marginal using Equation (11).

Table 3 reports the results of the simulation, giving the true values of the parameters, the average estimated value, the standard deviation of the estimates, the average of the reported standard errors of the estimate, the coverage of large-sample, symmetric confidence intervals and the rejection probability of a Wald test.

For the situation of no missing data, all estimators show little bias and have close to nominal coverage for confidence intervals. For binary outcomes, the estimator based on the joint model is about 30% more efficient (ratio of the variances approximately 1.3) and has slightly higher power. For the continuous outcomes, the gain is not as great, with the joint analysis being about 8% more efficient.

With data missing completely at random on the binary outcome, the availability of the (complete) continuous outcome in a joint analysis made the binary joint model even more efficient when compared with the separate analysis (over 40%) and the joint

Table 3 Simulation of joint and separate model fits to the probit/normal joint model

Data/estimator	True	Average	SD	Average SE	CI cov.	Pr{Reject}
No missing						
Joint – β_1^B	1	1.014	0.339	0.335	0.947	0.88
Joint – γ_1^B	0.707	0.720	0.238	0.237	0.949	0.88
Separate – γ_1^B	0.707	0.723	0.272	0.266	0.944	0.79
Joint – γ_1^C	2	1.988	0.282	0.268	0.935	1.00
Separate – γ_1^C	2	1.986	0.293	0.283	0.946	1.00
MCAR						
Joint – β_1^B	1	1.032	0.402	0.381	0.938	0.79
Joint – γ_1^B	0.707	0.729	0.279	0.270	0.944	0.79
Separate – γ_1^B	0.707	0.733	0.335	0.320	0.950	0.63
MAR (severe)						
Joint – β_1^B	1	1.018	0.369	0.363	0.943	0.84
Joint – γ_1^B	0.707	0.722	0.262	0.257	0.944	0.84
Separate – γ_1^B	0.707	0.644	0.334	0.329	0.945	0.50
MAR (mild)						
Joint – β_1^B	1	1.024	0.391	0.374	0.944	0.81
Joint – γ_1^B	0.707	0.726	0.274	0.265	0.944	0.81
Separate – γ_1^B	0.707	0.722	0.336	0.322	0.946	0.60

analysis lost only a small amount of power, whereas the power for the separate analysis dropped considerably.

In the severe missing at random scenario, the separate analysis was distinctly biased, had a much larger standard deviation and lost a significant amount of power. The joint analysis continued to perform well.

Finally, in the mild missing at random scenario, there was not evidence of bias for the separate analysis, and the efficiency loss was about the same as under MCAR.

Other authors have reported mixed results on efficiency gains, with a number of papers showing little or no gain.[32–34] These results here are more consistent with those of Gueorguieva and Sanacora[35] who show little improvement for a continuous outcome, but about a 20% decrease in standard errors for analysing an ordinal outcome in a joint model as opposed to alone.

Under all the scenarios, the average standard errors were a bit too small for all the estimators – slightly more so for the continuous outcomes. Confidence interval coverage was excellent.

9 Example: OAI

The OAI example is now reconsidered with joint modelling of the log transformation of the WOMAC score plus 1 (which is closer to approximate normality than the score itself) and the number of days of missed work in the past three months. Using subjects that were diagnosed with OA at enrollment as well as those who might develop OA gave a total of 2678 subjects for analysis. Three predictors are considered: age, sex and body mass index (BMI). Data are essentially complete for all variables except number of days of missed work, for which only 1605 subjects (about 60%) responded; the missing data are presumably due to those who did not work. It is easy to imagine that such data would not be missing completely at random, potentially biasing the fit of separate models.

A model of the form (19) is now fit, that is a linear regression of log of the WOMAC score plus 1 (LNWOM) on AGE, SEX and BMI and a Poisson regression of missed days of work (MISSW) on the same three predictors

$$\begin{aligned} \text{LNWOM}_i|b_i &\sim \text{indep. } \mathcal{N}(\mu_i^N, \sigma^2) \qquad (39)\\ \mu_i^N &= \beta_0^N + \beta_1^N \text{AGE}_i + \beta_2^N \text{SEX}_i + \beta_3^N \text{BMI}_i + \lambda b_i,\\ \text{MISSW}_i|b_i &\sim \text{indep. Poisson}(\mu_i^C) \qquad (40)\\ \log(\mu_i^C) &= \beta_0^C + \beta_1^C \text{AGE}_i + \beta_2^C \text{SEX}_i + \beta_3^C \text{BMI}_i + b_i, \end{aligned}$$

with $b_i \sim$ i.i.d. $\mathcal{N}(0, \sigma_b^2)$, using SAS Proc NLMIXED (SAS Institute, Cary, NC, USA). For comparison the models were also fit separately and without the random effect b_i.

Table 4 gives the coefficients and standard errors for the model fit. The shared random effect variance, σ_b^2, is very large and highly statistically significant, indicating strong evidence for a correlation between the two outcome types. Focussing on the regression coefficients, the estimates for the linear model are very similar for the separate and joint

Table 4 Parameter and standard error estimates (in parentheses) from joint and separate fits to the OAI data

Model	Parameter					
	β_{BMI}	β_{AGE}	β_{SEX}	$\log \sigma^2$	$\log \sigma_b^2$	$\log \lambda$
Linear LNWOM – separate	0.062	0.0082	0.28	0.39		
	(0.005)	(0.003)	(0.049)	(0.027)		
Linear LNWOM – joint	0.062	0.0082	0.28		2.79	
	(0.005)	(0.003)	(0.048)		(0.22)	
Poisson MISSW – separate	0.15	0.0011	−0.075			
	(0.012)	(0.079)	(0.13)			
Poisson MISSW – Joint	0.27	−0.51	−0.26	0.13	2.79	−1.95
	(0.051)	(0.31)	(0.48)	(0.073)	(0.22)	(0.15)

fits. This is not surprising given the simulation results since a) there is almost no missing data on the continuous outcome and b) the count data will contribute little information for the linear model fit.

However, for the count data model there are some differences. The effect of BMI nearly doubles, which is related to the fact that those with with no data on days missed from work tend to have larger values of BMI (average of 28.7 versus 28.2). The effect of sex also goes from small and not statistically significant to borderline statistically significant and potentially important: the estimated ratio of days missed from work is $\exp(-.26) = 0.77$, or 23% fewer missed days of work for females. As with BMI, sex is related to not having data on days missed from work (with 43% of females missing data compared with 34% of males). These both indicate that the use of joint models can lead to conclusions that are qualitatively different from fitting separate models.

10 Summary

Theoretical and numerical calculations have been used to illustrate the degree to which joint models for mixed outcome types can accommodate correlations and associations using shared and correlated random effects. These allow the modelling of correlated data – the usual scenario for multiple outcomes measured on the same subject. Joint models are scientifically necessary when the question of interest focuses on the joint behavior of multiple outcomes or when the association is of primary interest. Furthermore, it is argued that they are statistically advantageous in two aspects. First, they can increase efficiency, especially when data are missing on one of the outcomes, and especially when the outcome of primary interest is a 'low-information' outcome, such as a binary outcome. Secondly, they can protect against data that are missing at random, but not missing completely at random.

The use of shared and correlated random effects is natural for jointly normally distributed outcomes. In such a case, the inclusion of shared or correlated random effects modifies only the variance–covariance structure, leaving the mean structure and interpretation of regression coefficients unchanged. However, for nonnormal distributions, especially those like the Bernoulli and Poisson that have a natural tie between the mean

and variance, utilizing shared random effects can have additional consequences, which may or may not be advantageous.

These models can readily be fit using current software. Stata Ver 9.1 with the add-in module GLLAMM and SAS (SAS Institute), using its routine Proc NLMIXED is used.

References

1 Searle SR, Casella G, McCulloch CE. *Variance components*. Wiley, 1992.

2 McCulloch CE, Searle SR. *Generalized, linear and mixed models*. Wiley, 2000.

3 Ritz J, Spiegelman D. A note about the equivalence of conditional and marginal regression models. *Statistical Methods in Medical Research* 2004; **13**: 309–23.

4 Arminger G, Kusters U. Latent trait models with indicators of mixed measurement level. In Langeheine R, Rost J eds. *Latent trait and latent class models*. Plenum, 1988: 51–73.

5 Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process (Corr: V45 P1347; V47 P357). *Biometrics* 1988; **44**: 175–88.

6 Pawitan Y, Self S. Modeling disease marker processes in AIDS. *Journal of the American Statistical Association* 1993; **88**: 719–26.

7 Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* 1995; **90**: 27–37.

8 Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–9.

9 Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; **96**(455): 895–905.

10 Lin H, Turnbull B, McCulloch C, Slate E. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002; **93**: 1124–9.

11 Catalano PJ, Scharfstein DO, Ryan LM, Kimmel CA, Kimmel GL. Statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. *Teratology* 1993; **47**: 281–90.

12 Regan MM, Catalano PJ. Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. *Biometrics* 1999; **55**: 760–8.

13 Geys H, Regan MM, Catalano PJ, Molenberghs G. Two latent variable risk assessment approaches for mixed continuous and discrete outcomes from developmental toxicity data. *Journal of Agricultural, Biological, and Environmental Statistics* 2001; **6**(3): 340–55.

14 Yu ZF, Catalano PJ. Quantitative risk assessment for multivariate continuous outcomes with application to neurotoxicology: the bivariate case. *Biometrics* 2005; **61**(3): 757–66.

15 Fitzmaurice G, Laird N. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* 1995; **90**: 845–52.

16 Dunson D. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* 2000; **62**: 355–66.

17 Dunson D, Herring A. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics* 2005; **6**: 11–25.

18 Catalano P. Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine* 1997; **16**: 883–900.

19 Speiss M. Estimation of a two-equation panel model with mixed continuous and ordered categorical outcomes and missing data. *Applied Statistics* 2006; **55**: 525–38.

20 Sorensen J, Dilley J, London J, Okin R, Delucchi K, Phibbs C. Case management for substance abusers with HIV/AIDS: a randomized clinical trial. *The American Journal of Drug and Alcohol Abuse* 2003; **29**(1): 133–50.

21 Masson C, Sorensen J, Phibbs C, Okin R. Predictors of medical service utilization among individuals with co-occurring HIV infection and substance abuse disorders. *AIDS Care* 2004; **16**(6): 744–55.
22 Beck A, Steer R. *Manual for the revised Beck Depression Inventory*. Psychological Corporation, 1987.
23 Bellamy N. *WOMAC osteoarthritis user's guide*. Victoria Hospital, 1995.
24 Link TM, Steinbach L, Ghosh S, Ries M, Lane N, Majumdar S. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. *Radiology* 2003; **226**: 373–81.
25 Little R, Rubin D. *Statistical analysis with missing data*. Wiley, 2002.
26 Johnson NL, Kotz S, Balakrishnan N. *Continuous univariate distributions*, second edition. Wiley, 1995.
27 Qu Y, Williams GW, Beck GJ, Medendorp SV. Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics* 1992; **48**: 1095–102.
28 Chan KS, Ledolter J. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* 1995; **90**: 242–52.
29 McCulloch C, Lin H, Slate E, Turnbull B. Discovering subpopulation structure with latent class models. *Statistics in Medicine* 2002; **21**: 417–29.
30 Laird NM. Missing data in longitudinal studies. *Statistics in Medicine* 1988; **7**: 305–15.
31 Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 2005; **128**: 301–23.
32 Lesaffre E, Molenberghs G. Multivariate probit analysis: a neglected procedure in medical statistics. *Statistics in Medicine* 1991; **10**: 1391–403.
33 Fitzmaurice G, Laird N. Regression models for a mixed discrete and continuous responses with potentially missing values. *Biometrics* 1997; **53**: 110–222.
34 Gueorguieva R, Agresti A. A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* 2001; **96**(455): 1102–12.
35 Gueorguieva R, Sanacora G. Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine* 2006; **25**: 1307–22.