24

Joint Continuous and Discrete Responses

24.1 Introduction

Statistical problems where various outcomes of a mixed nature are observed have been around for about a half century and are rather common at present. Perhaps the most common situation, whether in psychometry, biometry, or other fields, is that of the joint occurrence of a continuous, often normally distributed, and a binary or ordinal outcome. Emphasis can be placed on the determination of the entire joint distribution of both outcomes, or on specific aspects, such as the association in general or correlation in particular between both outcomes.

For the problem sketched above, there broadly are three approaches. The first one postulates a marginal model for the binary outcome and then formulates a conditional model for the continuous outcome, given the categorical one. For the former, one can use logistic regression, whereas for the latter conditional normal models are a straightforward choice, i.e., a normal model with the categorical outcome used as a covariate (Tate 1954). The second family starts from the reverse factorization, combining a marginal model for the continuous outcome with a conditional one for the categorical outcome. Conditional models have been discussed by Cox and Wermuth (1992, 1994b), Krzanowski (1988), and Little and Schluchter (1985). Schafer (1997) presents a so-called *general location model* where a number of continuous and binary outcomes can be modeled together.

The third model family directly formulates a joint model for the two outcomes. In this context, one often starts from an bivariate continuous

variable, one component of which is explicitly observed and the other one observed in dichotomized, or generally discretized, version only (Tate 1955). Molenberghs, Geys, and Buyse (2001) presented a model based on a Plackett-Dale approach, where a bivariate Plackett distribution is assumed, of which one margin is directly observed and the other one only after dichotomization. General multivariate exponential family based models have been proposed by Prentice and Zhao (1991), Zhao, Prentice, and Self (1992), and Sammel, Ryan, and Legler (1997).

Of course, these developments have not been limited to bivariate joint outcomes. One can obviously extend these ideas and families to a multivariate continuous outcome and/or a multivariate categorical outcome. For the first and second families, one then starts from conditional and marginal multivariate normal and appropriately chosen multinomial models. Such a model within the first family has been formulated by Olkin and Tate (1961). Within the third family, models were formulated by Hannan and Tate (1965) and Cox (1974) for a multivariate normal with a univariate bivariate or discrete variable.

Apart from an extension from the bivariate to the multivariate case, one can introduce other hierarchies as well. For example, each of the outcomes may be measured repeatedly over time, and there could even be several repeated outcomes in both the continuous and the categorical subgroup, and then some of the approaches described in Chapter 25 can be used. A very specific hierarchy stems from clustered data, where a continuous and a categorical, or several of each, are observed for each member of a family, a household, a cluster, etc. For the specific context of developmental toxicity studies, often conducted in rats and mice, a number of developments have been made. An overview of such methods, together with developments for probit-normal and Plackett-Dale based models, was presented in Regan and Catalano (2002). Catalano and Ryan (1992) and Fitzmaurice and Laird (1995) propose models for a combined continuous and discrete outcome, but differ in the choice of which outcome to condition on the other one. Both use generalized estimating equations to allow for clustering. Catalano (1997) extended the model by Catalano and Ryan (1992) to accommodate ordinal variables.

Regan and Catalano (1999a) proposed a probit-type model to accommodate joint continuous and binary outcomes in a clustered data context, thus extending the correlated probit model for binary outcomes (Ochi and Prentice 1984) to incorporate continuous outcomes. Geys *et al* (2001) used a Plackett latent variable to the same effect, extending the bivariate version proposed by Molenberghs, Geys, and Buyse (2001). Estimation in such hierarchical joint models can be challenging. Regan and Catalano (1999a) proposed maximum likelihood, but considered GEE as an option too (Regan and Catalano 1999b). Geys *et al* (2001) made use of pseudo-likelihood. Ordinal extensions have been proposed in Regan and Catalano (2000).

It is clear that the literature on joint modeling of outcomes of various natures is diverse and growing. A broad ranging review of hierarchical models for joint continuous and discrete models can be found in Regan and Catalano (2002). In this chapter, we will focus on a few methods. We will emphasize the case of a continuous and a binary outcome as a basic paradigm (Section 24.2). In particular, a probit-normal formulation will be developed (Section 24.2.1), a Plackett-Dale approach (Section 24.2.2), and a bivariate generalized linear mixed model of a joint nature (Section 24.2.3). Hierarchical versions will be discussed in Section 24.3. Using data from an opthalmology study, used in the context of surrogate marker validation and introduced in Section 2.9, a concept also discussed in Section 21.3, the methods presented will be illustrated.

24.2 A Continuous and a Binary Endpoint

In this section, we start of with the bivariate, non-hierarchical, setting. Extensions to the fully hierarchical case are the topic of Section 24.

Two modeling strategies can be considered to accommodate mixed binary–continuous endpoints. Indeed, the joint distribution of a mixed continuous–discrete outcome vector can always be expressed as the product of the marginal distribution of one of the responses and the conditional distribution of the remaining response given the former response. One can choose either the continuous or the discrete outcome for the marginal model. The main problem with such approaches is that no easy expressions for the association between both endpoints are obtained. Therefore, we opt for a more symmetric treatment of the two outcome variables. We treat the case where the surrogate is binary and the true endpoint is continuous. The reverse case is entirely similar.

Let $\widetilde{S}_i$ be a latent variable of which S_i is the dichotomized version. In Section 24.2.1 we will describe a bivariate normal model for $\widetilde{S}_i$ and T_i, resulting in a probit-linear model for S_i and T_i. Section 24.2.2 presents an alternative formulation based on the bivariate Plackett (1965) density and resulting in a Plackett-Dale model.

24.2.1 A Probit-normal Formulation

In this formulation, we assume the following model:

$$T_i = \mu_T + \beta X_i + \varepsilon_{Ti}, \tag{24.1}$$

$$\widetilde{S}_i = \mu_S + \alpha X_i + \varepsilon_{Si}, \tag{24.2}$$

where μ_S and μ_T are fixed intercepts and α and β are the fixed effects of the treatment X on the surrogate and true endpoints respectively. Further,

ε_{Si} and ε_{Ti} are correlated error terms, assumed to satisfy:

$$\begin{pmatrix} \varepsilon_{Ti} \\ \varepsilon_{Si} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\rho\sigma}{\sqrt{1-\rho^2}} \\ & \frac{1}{1-\rho^2} \end{pmatrix}\right]. \tag{24.3}$$

Model (24.1)–(24.2) specifies a bivariate normal density. The variance of $\widetilde{S}_i$ is chosen for reasons that will be made clear in what follows. From this model, it is easily seen that the density of T_i is univariate normal with regression given in (24.1) and variance σ^2, implying that the parameters μ_T, β, and σ^2 can be estimated using linear regression software with response T_i and single covariate Z_i. Similarly, the conditional density of $\widetilde{S}_i$, given X_i and T_i is

$$\begin{aligned} \widetilde{S}_i \quad \sim \quad & N\left[\left(\mu_S - \frac{\rho}{\sigma\sqrt{1-\rho^2}}\mu_T\right) + \left(\alpha - \frac{\rho}{\sigma\sqrt{1-\rho^2}}\beta\right) X_i \right. \\ & \left. + \frac{\rho}{\sigma\sqrt{1-\rho^2}} T_i; 1\right], \end{aligned} \tag{24.4}$$

having unit variance and thus motivating our earlier choice for the covariance matrix of T_i and $\widetilde{S}_i$. Note that in Chapters 21 and 22 the marginal variances were set equal to one. In principle, these choices are equivalent, as long as no additional variance parameter for the latent variables is introduced. The corresponding probability

$$P(S_i = 1|T_i, X_i) = \Phi_1(\lambda_0 + \lambda_X X_i + \lambda_T T_i), \tag{24.5}$$

where

$$\lambda_0 = \mu_S - \frac{\rho}{\sigma\sqrt{1-\rho^2}}\mu_T, \tag{24.6}$$

$$\lambda_X = \alpha - \frac{\rho}{\sigma\sqrt{1-\rho^2}}\beta, \tag{24.7}$$

$$\lambda_T = \frac{\rho}{\sigma\sqrt{1-\rho^2}}, \tag{24.8}$$

and Φ_1 is the standard normal cumulative density function. Note that (24.5) implicitly defines the cutoff value for the dichotomized version. The λ parameters can be found by fitting model (24.5) to S_i with covariates X_i and T_i. This can be done with standard logistic regression software if it allows to specify the probit rather than the logit link, such as the LOGISTIC and GENMOD procedures in SAS. Given the parameters from the linear regression on T_i (μ_T, β, and σ^2) and the probit regression on S_i (λ_0, λ_X, and λ_T), the parameters from the linear regression on $\widetilde{S}_i$ can now be obtained

from (24.6)–(24.8):

$$\mu_S = \lambda_0 + \lambda_T \mu_T, \tag{24.9}$$

$$\alpha = \lambda_Z + \lambda_X \beta, \tag{24.10}$$

$$\rho^2 = \frac{\lambda_T^2 \sigma^2}{1 + \lambda_T^2 \sigma^2}. \tag{24.11}$$

The asymptotic covariance matrix of the parameters (μ_T, β) can be found from standard linear regression output. The variance of $\widehat{\sigma}^2$ equals $2\sigma^4/N$. The asymptotic covariance of $(\widehat{\lambda}_0, \widehat{\lambda}_X, \widehat{\lambda}_T)$ follows from logistic (probit) regression output. These three statements yield the covariance matrix of the six parameters upon noting that it is block-diagonal. To derive the asymptotic covariance of (μ_S, α, ρ) it suffices to calculate the derivatives of (24.9)–(24.11) with respect to the six original parameters and apply the delta method. They are:

$$\frac{\partial(\mu_S, \alpha, \rho)}{\partial(\mu_T, \beta, \sigma^2, \lambda_0, \lambda_X, \lambda_T)} = \begin{pmatrix} \lambda_T & 0 & 0 & 1 & 0 & \mu_T \\ 0 & \lambda_T & 0 & 0 & 1 & \beta \\ 0 & 0 & h_1 & 0 & 0 & h_2 \end{pmatrix},$$

where

$$h_1 = \frac{1}{2\rho} \frac{\lambda_T^2}{(1 + \lambda_T^2 \sigma^2)^2},$$

$$h_2 = \frac{1}{2\rho} \frac{2\lambda_T \sigma^2}{(1 + \lambda_T^2 \sigma^2)^2}.$$

Molenberghs, Geys, and Buyse (2001) developed a program in GAUSS that performs the joint estimation directly by maximizing the likelihood based on contributions (24.1) and (24.5).

24.2.2 A Plackett-Dale Formulation

Assume that the cumulative distributions of S_i and T_i are given by F_{S_i} and F_{T_i}. The joint cumulative distribution of both these quantities has been studied by Plackett (1965) and is discussed for the bivariate binary and ordinal cases in Section 7.7:

$$F_{T_i,S_i} = \begin{cases} \dfrac{1 + (F_{T_i} + F_{S_i})(\psi_i - 1) - C(F_{T_i}, F_{S_i}, \psi_i)}{2(\psi_i - 1)} & \text{if } \psi_i \neq 1, \\ F_{T_i} F_{S_i} & \text{if } \psi_i = 1, \end{cases}$$

where ψ_i, $C(\cdot)$, F_{T_i}, and F_{S_i} take the roles of ψ, $S(\cdot)$, μ_{1+}, and μ_{+1} in (7.40), respectively.

We can now derive a bivariate Plackett "density" function $G_i(t,s)$ for mixed continuous- binary outcomes. Suppose the success probability for S_i is denoted by π_i, then we can define $G_i(t,s)$ by specifying $G_i(t,0)$ and $G_i(t,1)$ such that they sum to $f_{T_i}(t)$. If we define

$$G_i(t,0) = \frac{\partial F_{T_i,S_i}(t,0)}{\partial t},$$

then this leads to specifying G_i by:

$$G_i(t,0) = \begin{cases} \frac{f_{T_i}(t)}{2}\left(1 - \frac{1+F_{T_i}(t)(\psi_i-1)-F_{S_i}(s)(\psi_i+1)}{C\left(F_{T_i},1-\pi_i,\psi_i\right)}\right) & \text{if } \psi_i \neq 1, \\ f_{T_i}(t)(1-\pi_i) & \text{if } \psi_i = 1, \end{cases} \tag{24.12}$$

and

$$G_i(t,1) = f_{T_i}(t) - G_i(t,0). \tag{24.13}$$

In this formulation we assume $T_i \sim N(\mu_i, \sigma^2)$, with $\mu_i = \mu_T + \beta X_i$ and $\text{logit}(\pi_i) = \mu_S + \alpha X_i$ with similar notation as in the probit case. The global odds ratio is assumed to be constant, but this is obviously open to extension. If we write

$$\boldsymbol{\theta}_i = \begin{pmatrix} \mu_i \\ \sigma^2 \\ \pi_i \\ \psi \end{pmatrix} \text{ and } \boldsymbol{\eta}_i = \begin{pmatrix} \mu_i \\ \ln(\sigma^2) \\ \text{logit}(\pi_i) \\ \ln(\psi) \end{pmatrix},$$

estimates of the regression parameters $\boldsymbol{\nu} = (\boldsymbol{\mu}, \beta, \alpha, \ln\sigma^2, \ln\psi)$ are easily obtained by solving the estimating equations $\boldsymbol{U}(\boldsymbol{\nu}) = \boldsymbol{0}$, using a Newton-Raphson iteration scheme, where $\boldsymbol{U}(\boldsymbol{\nu})$ is given by:

$$\sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\nu}}\right)' \left\{\left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\theta}_i}\right)'\right\}^{-1} \left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \ln G_i(t_i, s_i)\right).$$

24.2.3 A Generalized Linear Mixed Model Formulation

The developments in Section 8.8, where a linearization based marginal model has been presented, and in Chapter 14, where generalized linear mixed models have been introduced, can now be adapted to the present setting as well. In fact, it is useful to start from the formulation in Section 22.4, where both random effects and serial correlation have been allowed for. Expression (22.9) provides a general formulation, and (22.10) is specific for a random-effects logistic regression for repeated measures with serial, or residual, correlation. It is straightforward to consider this

framework in situations where various outcomes of a different nature are observed. In general, we merely have to write, as before,

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \tag{24.14}$$

where

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i) = \boldsymbol{h}(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i). \tag{24.15}$$

As usual, we assume $\boldsymbol{b}_i \sim N(\boldsymbol{0}, D)$. The key relaxing assumption is that the components of the inverse link functions $\boldsymbol{h}$ are allowed to change with the nature of the various outcomes in $\boldsymbol{Y}_i$. The variance of $\boldsymbol{\varepsilon}_i$ depends on the mean-variance links of the various outcomes, and can contain, in addition, a correlation matrix $R_i(\boldsymbol{\alpha})$ and overdispersion parameters ϕ_i. When there are no random effects in (24.15) a marginal model is obtained, as in Section 8.8. We will refer to this as a *marginal generalized linear models* (MGLM) approach. Reversely, assuming there are no residual correlations in $R_i(\boldsymbol{\alpha})$, a conditional independence model or purely random effects model results, which is still denoted by GLMM.

Using straightforward derivations, a general first-order approximate expression for the variance-covariance matrix of $\boldsymbol{Y}_i$ is:

$$V_i = \mathrm{Var}(\boldsymbol{Y}_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Sigma_i. \tag{24.16}$$

Here,

$$\Delta_i = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)\Bigg|_{\boldsymbol{b}_i = \boldsymbol{0}},$$

and

$$\Sigma_i \simeq \Xi_i^{1/2} A_i^{1/2} R_i(\boldsymbol{\alpha}) A_i^{1/2} \Xi_i^{1/2},$$

with A_i a diagonal matrix containing the variances following from the generalized linear model specification of Y_{ij} given the random effects $\boldsymbol{b}_i = \boldsymbol{0}$, i.e., with diagonal elements $v(\mu_{ij}|\boldsymbol{b}_i = \boldsymbol{0})$. Likewise Ξ_i is a diagonal matrix with the overdispersion parameters along the diagonal. When an outcome component is normally distributed, the overdispersion parameter is σ_i^2 and the variance function is 1. For a binary outcome with logit link, we obtain

$$\mu_{ij}(\boldsymbol{b}_i = \boldsymbol{0})[1 - \mu_{ij}(\boldsymbol{b}_i = \boldsymbol{0})].$$

The evaluation under $\boldsymbol{b}_i = \boldsymbol{0}$ derives from a Taylor series expansion of the mean components around $\boldsymbol{b}_i = \boldsymbol{0}$.

When an exponential family specification is used for all components, with canonical link, $\Delta_i = A_i$ and we can write:

$$V_i = \mathrm{Var}(\boldsymbol{Y}_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Xi_i^{1/2} \Delta_i^{1/2} R_i(\boldsymbol{\alpha}) \Delta_i^{1/2} \Xi_i^{1/2}. \tag{24.17}$$

Under conditional independence R_i vanishes and

$$V_i = \mathrm{Var}(\boldsymbol{Y}_i) = \Delta_i Z_i D Z_i' \Delta_i' + \Xi_i^{1/2} \Delta_i \Xi_i^{1/2}. \tag{24.18}$$

For the setting already considered in Sections 24.2.1 and 24.2.2, a suitable version of (24.14) is:

$$\begin{pmatrix} S_i \\ T_i \end{pmatrix} = \begin{pmatrix} \mu_S + \lambda b_i + \alpha X_i \\ \dfrac{\exp[\mu_T + b_i + \beta X_i]}{1 + \exp[\mu_T + b_i + \beta X_i]} \end{pmatrix} + \begin{pmatrix} \varepsilon_{Si} \\ \varepsilon_{Ti} \end{pmatrix}. \tag{24.19}$$

Note that we have included a scale parameter λ in the continuous component of an otherwise random-intercept model, given the continuous and binary outcome are measured on different scales. In this case,

$$Z_i = \begin{pmatrix} \lambda \\ 1 \end{pmatrix}, \qquad \Delta_i = \begin{pmatrix} 1 & 0 \\ 0 & v_{i2} \end{pmatrix}, \qquad \Phi = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix},$$

with $v_{i2} = \mu_{i2}(\boldsymbol{b}_i = \mathbf{0})[1 - \mu_{i2}(\boldsymbol{b}_i = \mathbf{0})]$. Further, let ρ be the correlation between ε_{Si} and ε_{Ti}. Note that Z_i is not a design matrix in the strict sense, since it contains an unknown parameter. Nevertheless, it is useful to consider this decomposition.

This implies that (24.16) becomes

$$\begin{aligned} V_i &= \begin{pmatrix} \lambda^2 & v_{i2}\lambda \\ v_{i2}\lambda & v_{i2}^2 \end{pmatrix} \tau^2 + \begin{pmatrix} \sigma^2 & \rho\sigma\sqrt{v_{i2}} \\ \rho\sigma\sqrt{v_{i2}} & v_{i2} \end{pmatrix} \\ &= \begin{pmatrix} \lambda^2\tau^2 + \sigma^2 & v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}} \\ v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}} & v_{i2}^2\tau^2 + v_{i2} \end{pmatrix}. \end{aligned} \tag{24.20}$$

The approximate marginal correlation function derived thereof equals:

$$\rho(\boldsymbol{\beta}) = \frac{v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}}}{\sqrt{\lambda^2\tau^2 + \sigma^2}\sqrt{v_{i2}^2\tau^2 + v_{i2}}}. \tag{24.21}$$

Obviously, (24.21) depends on the fixed effects through v_{i2}. In the special case of no random effects, the model can be written as:

$$\begin{pmatrix} S_i \\ T_i \end{pmatrix} = \begin{pmatrix} \mu_S + \alpha X_i \\ \dfrac{\exp(\mu_T + \beta X_i)}{1 + \exp(\mu_T + \beta X_i)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{Si} \\ \varepsilon_{Ti} \end{pmatrix}, \tag{24.22}$$

and (24.21) simply reduces to ρ, by virtue of its fully marginal specification. Under conditional independence, ρ in (24.20) satisfies $\rho \equiv 0$ and (24.21) reduces to

$$\rho(\boldsymbol{\beta}) = \frac{v_{i2}\lambda\tau^2}{\sqrt{\lambda^2\tau^2 + \sigma^2}\sqrt{v_{i2}^2\tau^2 + v_{i2}}}, \tag{24.23}$$

somewhat simpler but still a function of the fixed effects.

In case both endpoints are binary, the counterpart to (24.21) is

$$\rho(\boldsymbol{\beta}) = \frac{v_{i1}v_{i2}\tau^2 + \rho\sigma\sqrt{v_{i1}v_{i2}}}{\sqrt{v_{i1}^2\tau^2 + v_{i1}}\sqrt{v_{i2}^2\tau^2 + v_{i2}}}, \tag{24.24}$$

with again a constant correlation ρ when there are no random effects and, when there is no residual correlation:

$$\rho(\boldsymbol{\beta}) = \frac{v_{i1}v_{i2}\tau^2}{\sqrt{v_{i1}^2\tau^2 + v_{i1}}\sqrt{v_{i2}^2\tau^2 + v_{i2}}}, \tag{24.25}$$

Of course, the above calculations can be performed with ease for general random effects design matrices Z_i and for more than two components, of arbitrary nature and not just continuous and binary. This is useful, for example, for a fully hierarchical specification such as in Section 24.3.

In the general model, no full joint distribution needs to be specified, even when we assume the first one to be normally distributed, and the second one to be Bernoulli distributed. We still can leave the specification of the joint moments to the second one, by way of the marginal correlation. A full joint specification would need full bivariate model specification, conditional upon the random effects.

Under conditional independence, the specification of the outcome distributions conditional upon the random effects, together with the normality assumptions made about the random effects, fully specifies the joint distribution.

24.3 Hierarchical Joint Models

In the previous section, bivariate models have been discussed for the joint analysis of a continuous and a binary outcome. The focus was placed on a probit-normal and a Plackett-Dale formulation, next to the generalized linear mixed model framework, which can be used to flexibly derive marginal as well as random-effects models. Of course, joint outcomes can be measured repeatedly over time, or might be observed within a hierarchical context. In Section 24.3.1, a two-stage approach is presented, whereas Section 24.3.2 discusses fully hierarchical models.

24.3.1 Two-stage Analysis

In this section, we retain the setting of a binary and a continuous endpoint, measured within a hierarchical setting. Molenberghs, Geys, and Buyse (2001) used this approach in the context of surrogate marker evaluation. Let $\widetilde{S}_{ij}$ be a latent variable of which S_{ij} is a dichotomized version. One

option is to consider a two-step analysis. Assume that subject j is measured within trial i. For repeated measures, j would refer to time and i to subject.

At the first step, we can assume the following model:

$$\begin{aligned} \widetilde{S}_{ij} &= \mu_{S_i} + \alpha_i X_{ij} + \varepsilon_{S_{ij}}, \\ T_{ij} &= \mu_{T_i} + \beta_i X_{ij} + \varepsilon_{T_{ij}}, \end{aligned}$$

where α_i and β_i are study-specific effects of treatment X on the endpoints in trial i, μ_{S_i} and μ_{T_i} are trial-specific intercepts, and ε_{S_i} and ε_{T_i} are correlated error terms, assumed to be mean-zero normally distributed with covariance matrix

$$\Sigma = \begin{pmatrix} \frac{1}{(1-\rho^2)} & \frac{\rho\sigma}{\sqrt{1-\rho^2}} \\ \frac{\rho\sigma}{\sqrt{1-\rho^2}} & \sigma^2 \end{pmatrix}.$$

In short, we use the probit formulation, described in Section 24.2.1. Due to the replication at the study level, we can impose a distribution on the study-specific parameters. At the second stage we assume

$$\begin{pmatrix} \mu_{S_i} \\ \mu_{T_i} \\ \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_T \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} m_{S_i} \\ m_{T_i} \\ a_i \\ b_i \end{pmatrix} \tag{24.26}$$

where the second term on the right hand side of (24.26) is assumed to follow a zero-mean normal distribution with dispersion matrix D.

24.3.2 Fully Hierarchical Modeling

We first indicate how the probit-normal and Plackett-Dale models can be generalized to the hierarchical setting. Ample detail can be found in Geys *et al* (2001) and Regan and Catalano (2002). Next, the generalized linear mixed model case will be considered.

24.3.2.1 A Probit-normal Formulation

The model of Section 24.2.1 can be seen as the basis for this model. Whereas Model (24.1)–(24.2) applies to one continuous and one binary outcome, we could equally well consider multiple copies of each and then assume that the resulting stochastic vector, composed of directly observed and latent outcomes, is normally distributed.

Although this approach is natural and appealing, the problem is the handling of potentially high dimensional probits, and several authors have considered this problem in detail. Regan and Catalano (1999a) introduced a mixed-outcome probit model that extends a correlated probit model for

binary outcomes (Ochi and Prentice 1984) to incorporate continuous outcomes. These authors consider exchangeability among the continuous outcomes, among the binary outcomes, and between the continuous and binary outcomes.

Regan and Catalano (1999b) avoided fully specifying the joint distribution of the n_i bivariate outcomes on related subjects within unit i by specifying only the marginal distribution of the bivariate outcomes and applying generalized estimating equations to take correlation into account. Precisely, they fully model the bivariate outcomes for a subject and then apply GEE to accommodate for the correlations between subjects within unit i.

24.3.2.2 A Plackett-Dale Approach

Likewise, the Plackett-Dale model of Section 24.2.2 can be embedded in a hierarchical setting. Geys *et al* (2001) applied marginal pseudo-likelihood ideas (Chapter 9)

In Section 24.2.2, a bivariate density-distribution was defined for a joint continuous and binary outcome, by means of (24.12)–(24.13). In principle, a $2n_i$-dimensional Plackett-Dale model needs to be specified. Alternatively, progress can be made by solely specifying the bivariate outcomes, just as before, and assembling them into a (log) pseudo-likelihood function:

$$p\ell = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \ln G_{ij}(t_{ij}, s_{ij}), \tag{24.27}$$

where T_{ij} is the continuous outcomes for subject j within unit (study, trial, center,...) i and S_{ij} is the binary one. Thus, with this particular choice of pseudo-likelihood function, the longitudinal part of the correlation structure is left unspecified. Of course, alternative pseudo-likelihood functions can be used as well, depending on which parameters are needed to formulate answers to scientific questions. Sometimes, the correlation structure between outcomes on different subjects within the same unit can be of interest, calling for other types of pseudo-likelihood function. Parameter and precision estimation based on (24.27) is straightforward, given the developments in Chapter 9, in particular Section 9.4.

24.3.2.3 A Generalized Linear Mixed Model Formulation

The developments in Section 24.2.3 extend straightforwardly to the hierarchical case, including repeated measures, meta-analyses, clustered data, correlated data, etc. In fact, Model (24.14) is sufficiently general to generate marginal and random-effects models for such settings. The fixed and random effects structures can be formulated sufficiently generally so as to cover all of these settings. Of course, when parameters are shared between models for outcomes of different types, care has to be taken to ensure the

models are meaningful. For example, inflation factors might have to be used to share random effects across binary and continuous outcomes, exactly as the parameter λ in (24.19).

Correlations follow in a straightforward fashion when purely marginal versions are used. When random effects are involved, correlation structures can be derived from (24.16) or specific forms derived thereof.

24.4 Age Related Macular Degeneration Trial

In the Age Related Macular Degeneration Study, introduced in Section 2.9, the mixed discrete-continuous case is encountered in data from a simple yet real situation. Indeed, visual acuity is assessed in terms of number of letters read, which can be treated as continuous. The dichotomization in terms of at least 2 or 3 lines of vision lost at 6 and 12 months, respectively, is a binary outcome.

In Section 24.4.1, a number of bivariate marginal analyses are presented, with bivariate random-effects analyses discussed in Section 24.4.2. Hierarchical analyses, based on including center as a hierarchy defining variable on the one hand, and repeated measures on each of the binary and continuous outcomes on the other hand, are presented in Section 24.4.3.

24.4.1 Bivariate Marginal Analyses

First, we consider dichotomized visual acuity at 6 months as the surrogate and (continuous) visual acuity at 12 months as the true endpoint. Dichotomization is achieved by setting a binary variable to 1 if visual acuity at 6 months is larger than the value at baseline and to 0 otherwise. We consider a probit-normal model as in Section 24.2.1, a Plackett-Dale model as in Section 24.2.2, and a GLM-based marginal model as in Section 24.2.3. Of course, the roles of S_i and T_i are reversed in the corresponding equations, as here the surrogate is assumed binary while the true outcome was binary in the earlier sections. For the latter model, both a logit as well as a probit link is considered for the MGLM. PQL is used as approximation method. For the Plackett-Dale model, a logit link is employed for the true endpoint. Parameter estimates (standard errors) are displayed in Table 24.1.

The correlation between both endpoints is estimated as $\widehat{\rho} = 0.74$ under the probit model. This parameter is of direct interest in surrogate marker evaluation since it captures the so-called adjusted association (Buyse and Molenberghs 1998) or individual-level association (Buyse *et al* 2000, Molenberghs, Geys, and Buyse 2001). It also justifies the use of a joint model for both endpoints, rather than considering them separately. This parameter is estimated very precisely and there is apparently a strong correlation between both endpoints. Now, the corresponding correlation under the GLM

TABLE 24.1. *Age Related Macular Degeneration Trial. Bivariate marginal analyses with a binary surrogate and a continuous true endpoint.*

		probit-	Plackett	MGLM	
Effect	Par.	normal	-Dale	logit	probit
		Binary surrogate endpoint			
Intercept	μ_S	0.64(0.20)	0.74(0.19)	1.25(0.24)	0.76(0.14)
Treatm. eff.	α	0.39(0.28)	0.45(0.30)	0.40(0.38)	0.23(0.21)
Overdis. par.	ϕ			1.01(0.10)	1.01(0.10)
		Continuous true endpoint			
Intercept	μ_T	11.04(1.57)	10.89(1.56)	11.04(1.58)	11.04(1.58)
Treatm. eff.	β	4.12(2.32)	4.02(2.32)	4.12(2.33)	4.12(2.33)
Standard dev.	σ_T	15.95(0.82)	16.04(0.81)		
Variance	σ_T^2	254.4(26.2)	257.3(26.0)	257.0(26.5)	257.0(26.5)
		Association			
Correlation	ρ	0.74(0.05)		0.62(0.05)	0.62(0.05)
Log odds r.	$\ln \psi$		2.85(0.37)		
Odds r.	ψ		17.29(6.40)		

is quite a bit lower. Although, due to the use of PQL, there typically is downward bias in the parameter estimates, a more important reason for the difference is that the probit model features the correlation between a pair of *latent* variables, whereas the GLM captures the correlation between the observable outcomes. The Plackett-Dale model, of course, is based on the use of the odds ratio rather than the correlation as association parameter. For the binary endpoint, the treatment effect parameters differ somewhat, with the differences in the intercepts a bit larger. The parameter estimates for the continuous endpoint agree much closer.

Let us now switch to the situation of continuous visual acuity at 6 months as a surrogate for the binary indicator for loss of at least 3 lines of vision lost at one year. The same models as in Table 24.1 are considered here too, with of course the roles of the continuous and binary endpoints reversed. Parameter estimates (standard errors) are given in Table 24.2. Qualitative conclusions agree very closely with their counterparts for the earlier analyses, although there are some quantitative differences. With the probit model, the correlation is $\widehat{\rho} = 0.81$, but again, for the GLM-based models they are quite a bit smaller, underscoring once more that the two correlation parameters are not really directly comparable, as the probit (and also Dale) versions are describing the correlation of the underlying bivariate latent variable. With the Plackett-Dale model, the odds ratio is estimated to be $\widehat{\psi} = 16.93$. As in Table 24.1, parameter estimates across models agree

TABLE 24.2. *Age Related Macular Degeneration Trial. Bivariate marginal analyses with a continuous surrogate and a binary true endpoint.*

		Probit-	Plackett	MGLM	
Effect	Par.	normal	-Dale	logit	probit
Continuous surrogate endpoint					
Intercept	μ_S	5.53(1.26)	5.89(1.24)	5.53(1.27)	5.53(1.27)
Treatm. eff.	α	2.83(1.87)	2.72(1.84)	2.83(1.87)	2.83(1.87)
Standard dev.	σ_S	12.80(0.66)	12.90(0.65)		
Variance	σ_S^2	163.8(16.9)	166.4(16.8)	165.7(17.1)	165.7(17.1)
Binary true endpoint					
Intercept	μ_T	-0.36(0.21)	-0.36(0.19)	-0.50(0.20)	-0.31(0.13)
Treatm. eff.	β	0.60(0.30)	0.58(0.28)	0.66(0.30)	0.41(0.19)
Overdis. par.	ϕ			1.01(0.10)	1.01(0.10)
Association					
Correlation	ρ	0.81(0.04)		0.62(0.04)	0.62(0.04)
Log odds r.	$\ln \psi$		2.83(0.29)		
Odds r.	ψ		16.93(4.91)		

fairly closely, but the agreement is better for the continuous endpoint than for the binary one.

Of course, one could also analyze both endpoints as binary, or both endpoints as continuous. Although not the theme of the chapter, it is useful to do so for the sake of comparison. In the first case, a standard probit or Dale model (Chapter 7) could be used. In the second case, a bivariate normal is the obvious choice. Let us first focus on the situation of two binary outcomes. Buyse and Molenberghs (1998) analyzed both binary endpoints using the Dale model with logit links and obtained an odds ratio of $\widehat{\psi} = 18.53$. Table 24.3 presents five different analyses of the pair of binary outcomes. First, the Dale model is fitted with both logit and probit links. Second, a marginal linearization based model with correlated error terms (Section 8.8 is considered, again with logit and probit links. Third, a bivariate probit model is fitted. Table 24.3 organizes the models by link functions, so that similarities and differences between parameter estimates become more apparent.

Even more so than in the heterogenous outcome cases, there is close agreement between the intercept and treatment effect parameter estimates for the logit and probit models, respectively. At the same time, there is agreement between the association measures as far as they are comparable, but once again the probit based correlation is quite a bit higher than the GLM-based correlation, for reasons explained above.

TABLE 24.3. *Age Related Macular Degeneration Trial. Bivariate marginal analyses with binary endpoints, based on the Dale model (probit and logit links), the bivariate probit model, and a marginal joint GLM (logit and probit links).*

		Logit links	
Effect	Parameter	Dale	MGLM
Surrogate endpoint			
Intercept	μ_S	-0.54(0.20)	-0.54(0.21)
Treatm. eff.	α	0.70(0.30)	0.70(0.30)
Overdis. par.	ϕ		1.01(0.10)
True endpoint			
Intercept	μ_T	-0.50(0.20)	-0.50(0.20)
Treatm. eff.	β	0.66(0.30)	0.66(0.30)
Overdis. par.	ϕ		1.01(0.10)
Association			
Correlation	ρ		0.62(0.05)
Log odds r.	$\ln \psi$	2.92(0.38)	
Odds r.	ψ	18.54(7.05)	

		Probit links		
Effect	Parameter	biv. probit	Dale	MGLM
Surrogate endpoint				
Intercept	μ_S	-0.34(0.13)	-0.33(0.13)	-0.33(0.13)
Treatm. eff.	α	0.44(0.18)	0.44(0.18)	0.44(0.19)
Overdis. par.	ϕ			1.01(0.10)
True endpoint				
Intercept	μ_T	-0.31(0.13)	-0.31(0.13)	-0.31(0.13)
Treatm. eff.	β	0.41(0.18)	0.41(0.18)	0.41(0.19)
Overdis. par.	ϕ			1.01(0.10)
Association				
Correlation	ρ	0.83(0.05)		0.62(0.05)
Log odds r.	$\ln \psi$		2.92(0.38)	
Odds r.	ψ		18.54(7.05)	

Finally, both outcomes can be considered continuous. Then, the counterparts of all models in Tables 24.1–24.3 collapse to a bivariate normal model, and so does the output obtained from virtually all relevant software tools, such as the SAS procedures MIXED, NLMIXED, and GLIMMIX. Results are presented in Table 24.4. The correlation obtained here is 0.75. Note that this is closer to the bivariate probit and probit-normal models than to the GLM one. Indeed, we now have a bivariate continuous outcome, which

TABLE 24.4. *Age Related Macular Degeneration Trial. Bivariate marginal analyses with continuous endpoints, using a bivariate normal model.*

Effect	Par.	Estimate (s.e.)
Surrogate endpoint		
Intercept	μ_S	5.53(1.27)
Treatm. eff.	α	2.83(1.87)
Standard dev.	σ_S	12.87(0.66)
Variance	σ_S^2	165.7(17.1)
True endpoint		
Intercept	μ_T	11.04(1.58)
Treatm. eff.	β	4.12(2.33)
Standard dev.	σ_T	16.03(0.83)
Variance	σ_T^2	257.0(26.5)
Association		
Correlation	ρ	0.75(0.03)

is more informative than a pair of binary outcomes or the joint occurrence of a binary and a continuous outcome. Nevertheless, in all situations do the probit and probit-normal models attempt to describe the association of the underlying pair of normal outcomes, whether or not they are directly observed.

Generally note that, when continuous or binary outcome results are compared across Tables 24.1–24.4, whether from a heterogeneous or homogenous model, there is reasonably close agreement, especially within a model family (probit-normal, Plackett-Dale, GLM based), and especially for treatment effects and association parameters.

24.4.2 Bivariate Random-effects Analyses

Although all models above are of a marginal type, we can also consider random-effects models. So far, we have considered marginal versions of (24.14), denoted by MGLM, but we will now switch to conditional independence model (24.19) with a scaled random intercept, for the case of a binary and a continuous outcome, a classical random-intercepts logistic regression model when both outcomes are binary, and a random-intercept linear mixed-effects model for continuous outcomes. Results are presented in Table 24.5.

Comparing the continuous-binary case with the results in Table 24.2, it is clear that fixed effects in the Gaussian model roughly remain the same, but the fixed effects for the binary outcome are larger, in agreement with the results in Chapter 16. A similar inflation is seen in the binary-binary

case, at least when numerical integration is used. For PQL, the bias is severe and the parameter estimates are hardly larger than their marginal counterparts in Table 24.3. Given the estimate of the random-intercept variance ($\hat{\tau}^2 = 14.51$) and (16.3), the correspondence between the random-effects parameters and their marginal counterparts in Table 24.3 would be 2.45. Comparing the corresponding estimates yields factors of roughly 2.62. In line with general results about the linear mixed model (Chapter 4), the estimates in the last column are very close to those in Table 24.4, even though the assumption of a constant variance, made here, may be somewhat too simplistic, given the variances in Table 24.4 are quite a bit different.

As before, the correlation between both endpoints is of interest. With the exception of the continuous-continuous case, it is somewhat less straightforward to derive. For the continuous-binary case, we can make use of (24.23) and for the binary-binary case, (24.24) is the proper choice. Clearly, the correlation is different between both treatment arms now, given the dependence of the correlation function on the fixed effects. However, in this case, the difference is negligible. However, the poverty of the PQL approximation is shown, not only in the fixed effect and variance component estimates, but in the correlation parameters as well. For the others, irrespective on the nature of the outcomes, the results are very close to their marginal counterparts in Tables 24.1–24.4, which is reassuring.

It is worthwhile to note that the parameters in the continuous-binary case are identifiable, but due to the non-linearity of the model, induced by the factor λ, care has to be taken in monitoring the convergence process. Having said this, the effect is most clearly seen on the binary outcome fixed effects, and not quite as much on the continuous outcome parameters.

24.4.3 Hierarchical Analyses

Let us now switch attention to the hierarchical case. First, let us observe that the trial is of the multicenter type. It is natural to consider the center in which the patients were treated as the unit of analysis. A total of 36 centers were thus available for analysis, with a number of individual patients per center ranging from 2 to 18. We analyze the situation where dichotomized visual acuity at 6 months acts as surrogate for the continuous visual acuity at 12 months. A two-stage approach is followed. Table 24.6 shows the parameter estimates for the hierarchical probit model (Section 24.3.1). Two versions are considered, with trial-specific treatment effects on the one hand (reduced model) and with trial-specific intercepts and treatment effects on the other hand (full model). The correlation, obtained from the full model, is similar to the ones obtained from the bivariate analyses. When the reduced model is employed, the correlation is quite a bit smaller.

Of course, also fully hierarchical models can be fitted. For example, the hierarchical probit or Plackett-Dale models can be used. Applications of these models can be found in Regan and Catalano (2002). Also the joint

TABLE 24.5. *Age Related Macular Degeneration Trial. Bivariate joint generalized linear mixed model analyses. Some models lead to a treatment-arm dependent correlation estimate, denoted by 'stand' for the standard arm and 'exp' for the experimental arm.*

Surrogate endpoint:		cont.	binary	binary	cont.
True endpoint:		binary	binary	binary	cont.
Estimation method:		PQL	Num. int.	PQL	ML
Effect	Par.				
Surrogate endpoint parameters					
Intercept	μ_S	5.53(1.26)	1.42(0.57)	-0.62(0.26)	5.53(1.42)
Treatm. eff.	α	2.83(1.86)	-1.84(0.82)	0.81(0.39)	2.83(2.11)
Standard dev.	σ_S	7.18(1.15)			
Variance	σ_S^2	51.59(16.55)			
Inflation	λ	-1.41(1.68)			
True endpoint parameters					
Intercept	μ_T	1.63(1.94)	1.31(0.56)	-0.57(0.26)	11.04(1.42)
Treatm. eff.	β	-2.72(3.15)	-1.73(0.81)	0.76(0.39)	4.12(2.11)
Common parameters, including association					
R.I. std.d.	τ	7.50(8.50)	3.81(0.69)	1.95(0.47)	12.41(0.76)
R.I. var.	τ^2	56.2(127.4)	14.51(5.28)	3.76(1.82)	154.0(18.8)
Res. st.d.	σ				7.43(0.38)
Res. var.	σ^2				55.1(15.7)
Correlation	ρ				0.74
Corr. (stand.)	$\rho[1]$	0.79	0.78	0.48	
Corr. (exp.)	$\rho[2]$	0.78	0.70	0.46	

generalized linear mixed effects model of Section 24.2.3 can be used for hierarchical analyses. Although we have focused so far on outcomes at 6 months and 1 year, we will now also consider the intermediate endpoints at 4 and 12 weeks as well. Thus, we have two repeated sequences of four components each, one binary, and one continuous. The binary outcomes are dichotomizations of the number of letters lost as negative *versus* non-negative. We consider on the one hand a marginal model, with fully unstructured 8×8 variance-covariance matrix and a conditional independence random-intercepts model on the other hand. Parameter estimates are presented in Table 24.7.

Once more, the relationship between the fixed effects is in line with expectation. For the continuous outcome sequence, they are virtually equal. For the binary outcome, the ratios vary between 1.55 and 1.98, with an average of 1.80, whereas (16.3) predicts a ratio of 1.80. Model parameters here are better identifiable than their counterparts from the bivariate models, even

TABLE 24.6. *Age Related Macular Degeneration Trial. Parameter estimates (standard errors) for the full and reduced two-stage fixed effects probit model.*

Effect	Parameter	Full	Reduced
	Surrogate endpoint		
Intercept	μ_S	1.46(0.68)	0.67(0.15)
Treatm. eff.	α	1.10(0.98)	1.75(0.69)
	True endpoint		
Intercept	μ_T	11.13(1.69)	11.82(1.00)
Treatm. eff.	β	4.40(2.94)	3.72(2.38)
Standard dev.	σ_T	11.43(0.60)	13.60(0.71)
Variance	σ_T^2	130.6(13.7)	185.0(19.3)
	Association		
Correlation	ρ	0.75(0.05)	0.66(0.07)

though care is still needed when selecting starting values. Every possible pair of outcomes in the marginal model has its own correlation coefficient (not shown), whereas in the random-effects model, they follow from the fixed effects and variance components, as was illustrated in Section 24.4.2, based on such expressions as (24.21), (24.23), (24.24), and (24.25).

24.5 Joint Models in SAS

We will present a program and selected output for the joint analysis of a continuous and a binary outcome, by means of the generalized linear mixed model and using the SAS procedure GLIMMIX. To create the bivariate outcome vectors, out of the continuous outcomes measured at 6 months (24 weeks) and 12 months (52 weeks), 'diff24' and 'diff52,' and their binary counterparts 'bindif24' and 'bindif52,' the following code can be used:

```
data armd77;
set armd7;
array x (2) diff24 diff52;
array y (2) bindif24 bindif52;
array z (2) bindh24 diff52;
array w (2) diff24 bindif52;
do j=1 to 2;
   visual=x(j);
   bindif=y(j);
   bincont=z(j);
   contbin=w(j);
   time=j;
```

TABLE 24.7. *Age Related Macular Degeneration Trial. Hierarchical models for joint longitudinal continuous and binary visual acuity sequences. For the marginal model-based and empirically corrected standard errors are presented.*

Effect	Parameter	Marginal	Random Int.
	Continuous sequence		
Intercept 4	β_{11}	-3.26(0.77;0.81)	-3.27(1.30)
Intercept 12	β_{21}	-4.62(1.14:1.07)	-4.62(1.29)
Intercept 24	β_{31}	-8.37(1.38;1.26)	-8.37(1.29)
Intercept 52	β_{41}	-15.16(1.72;1.64)	-15.16(1.29)
Treatment eff. 4	β_{12}	2.31(1.05;1.05)	2.38(1.76)
Treatment eff. 12	β_{22}	2.34(1.54;1.52)	2.34(1.76)
Treatment eff. 24	β_{32}	2.83(1.87;1.84)	2.83(1.76)
Treatment eff. 52	β_{42}	4.12(2.33;2.31)	4.12(1.76)
Res. st. deviation	σ		8.21(0.23)
Res. variance	σ^2		67.45(3.81)
Inflation	λ		-3.32(0.34)
	Binary sequence		
Intercept 4	β_{11}	-1.02(0.24;0.24)	-2.02(0.46)
Intercept 12	β_{21}	-0.91(0.24;0.24)	-1.81(0.45)
Intercept 24	β_{31}	-1.15(0.25;0.25)	-2.24(0.47)
Intercept 52	β_{41}	-1.65(0.29;0.29)	-3.11(0.52)
Treatment eff. 4	β_{12}	0.40(0.32;0.32)	0.66(0.59)
Treatment eff. 12	β_{22}	0.54(0.31:0.31)	0.93(0.58)
Treatment eff. 24	β_{32}	0.52(0.33;0.32)	0.88(0.60)
Treatment eff. 52	β_{42}	0.40(0.38;0.38)	0.62(0.64)
	Common parameters		
R.I. st. deviation	τ		2.66(0.29)
R.I. variance	τ^2		7.07(1.64)

```
   subject=_n_;
   output;
end;
run;
```

There are four new outcomes created, consisting of the two continuous outcomes ('visual'), the two binary outcomes ('bindif'), a binary surrogate followed by a continuous true outcome ('bincont'), and finally a continuous surrogate followed by a binary true outcome ('contbin').

Because we cannot uniformly specify the outcome distribution nor the link function, a special device has been created to this effect, i.e., the 'byobs=(·)' specification that can be used in both the 'link=' and the 'dist='

options. Practically, a variable needs to be created to specify the outcome distribution and link function for each observation in the set of data. For example, analyzing the ARMD data with a continuous, normally distributed, surrogate and a binary true endpoint, can be done by means of the 'dist=byobs(distcb)' option where 'distcb' is a variable denoting a Gaussian distribution for the first measurement of every subject and a Bernoulli one for the second. The procedure recognizes both a numerical indicator, with a proper map between distributions and numerical labels being provided in the manual (SAS Institute Inc. 2004), as well as a four-character label, by means of the first four characters of each distributions. All but the multinomial distribution can be used. The following code creates to distributional indicators, one for a continuous surrogate and a binary true endpoint ('distcb') and one for the reverse case ('distbc'). In addition, four link function indicators are created, referring to the identity link for the continuous outcome and then either a logit or a probit link for the binary outcome.

Code to create these indicators is

```
data armd77;
set armd77;
distcb='BINA';
if time=1 then distcb='GAUS';
distbc='BINA';
if time=2 then distbc='GAUS';
linkcb1='LOGI';
if time=1 then linkcb1='IDEN';
linkcb2='PROB';
if time=1 then linkcb2='IDEN';
linkbc1='LOGI';
if time=2 then linkbc1='IDEN';
linkbc2='PROB';
if time=2 then linkbc2='IDEN';
run;
```

The relevant variables for analysis, for the first 5 subjects, are

	s					b	c			l	l	l	l
	u			v	b	i	o	d	d	i	i	i	i
	b		t	i	i	n	n	i	i	n	n	n	n
	j	t	r	s	n	c	t	s	s	k	k	k	k
0	e	i	e	u	d	o	b	t	t	c	c	b	b
b	c	m	a	a	i	n	i	c	b	b	b	c	c
s	t	e	t	l	f	t	n	b	c	1	2	1	2
1	1	1	1	0	0	1	0	GAUS	BINA	IDEN	IDEN	LOGI	PROB
2	1	2	1	-10	0	-10	0	BINA	GAUS	LOGI	PROB	IDEN	IDEN

```
  3   2  1  2   -3  0    0   -3  GAUS  BINA  IDEN  IDEN  LOGI  PROB
  4   2  2  2    1  0    1    0  BINA  GAUS  LOGI  PROB  IDEN  IDEN

  5   3  1  1   -6  1    1   -6  GAUS  BINA  IDEN  IDEN  LOGI  PROB
  6   3  2  1  -17  1  -17    1  BINA  GAUS  LOGI  PROB  IDEN  IDEN

  7   4  1  2    8  0    0    8  GAUS  BINA  IDEN  IDEN  LOGI  PROB
  8   4  2  2    1  0    1    0  BINA  GAUS  LOGI  PROB  IDEN  IDEN

  9   5  1  2   -2  0    1   -2  GAUS  BINA  IDEN  IDEN  LOGI  PROB
 10   5  2  2   -2  0   -2    0  BINA  GAUS  LOGI  PROB  IDEN  IDEN
...
```

The variable 'contbin' is clearly made up of the first component of 'visual' and the second one of 'bindif.' For 'bincont,' a somewhat different definition is used for the surrogate, which is an indicator for whether letters are lost or gained, rather than an indicator for at least two lines lost. This definition was chosen in agreement with the choice made by Molenberghs, Geys, and Buyse (2001).

We can now use the program:

```
proc glimmix data=armd77 method=rspl;
class treat distcb subject;
nloptions maxiter=50 technique=newrap;
model contbin = distcb treat*distcb
                / noint dist=byobs(distcb) solution;
random _residual_ / subject=subject type=un r;
run;
```

Note that there is no link function specification in this program, implying that the default link is used. Equivalently, one could specify the option 'link=byobs(linkcb1),' which would produce exactly the same model. However, the advantage then is that the link functions chosen become very explicit. Changing the link variable to 'link=byobs(linkcb2),' the probit link would be chosen for the binary variable, while maintaining the identity link for the continuous variable. The variable 'distcb' is also used in the fixed-effects structure, through the MODEL statements. This means that a separate intercept (μ_S and μ_T, respectively) and a separate treatment effect (α and β, respectively) are included for each of the two outcomes. This could be done equally well by using the variable 'time' as a class variable, since both 'time' and 'distcb,' and in fact also the link function variables, contain the same information when used as class variable. The choice for 'distcb' is motivated by clarity of the output, where it will be made clear which parameters belong to the Gaussian outcome and which to the binary one.

Given that the outcomes are of a different nature, this is a very natural choice. By including 'noint' into the MODEL statement options, both in-

tercepts are directly shown, rather than as a main effect and a difference between both, which would be less meaningful. The NLOPTIONS statement is included to control convergence. In examples like this, in agreement with the comments made in Section 22.6, convergence can be an issue and the user may need to change such aspects as the iterative technique, the maximum number of iterations, and the convergence tolerance.

Because the response and link functions depend on the outcome, the 'Model Information' panel does not specify them individually but rather gives a generic indication:

Model Information

Response Distribution	Multivariate
Link Function	Multiple
Variance Function	Default

Let us now turn to the estimates of the covariance parameters.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	subject	165.69	17.0897
UN(2,1)	subject	8.0235	1.1105
UN(2,2)	subject	1.0106	0.1042

The parameter 'UN(1,1)' is the variance of the Gaussian outcome, the parameter 'UN(2,2)' is the variance of the binary outcome and as such merely is an overdispersion parameter. Finally, 'UN(2,1)' is the covariance between both. In our example, the correlation is of interest more than the covariance. Because it can be calculated without problem from the three parameters, and the standard error could be calculated from the asymptotic covariance matrix of the variance parameters, it is actually easy to obtain it directly, but changing the structure option for the covariance matrix in the RANDOM statement to 'type=unr' rather than the 'type=un' structure used above. Obviously, both parameterizations are equivalent. The above panel then changes to:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Corr(2,1)	subject	165.69	17.0897
Corr(3,1)	subject	1.0106	0.1042
Corr(3,2)	subject	0.6200	0.04489

Of course, the two variance parameters are the same as above, but the correlation estimate $\widehat{\rho} = 0.62$ is now presented directly. Two observations are worth making. First, the order of the parameters in both panels is different and, somewhat misleading, the double indices have changed from the intuitive (1,1), (2,1), and (2,2) coding to (2,1), (3,1), and (3,2). These would correspond to correlations in a 3×3 correlation matrix, but not to the situation we encounter. So we advise to be careful with these labels and cautiously map the 'type=unr' parameters to their counterparts coming from the 'type=un' structure, to avoid confusion.

Finally, the fixed effects parameters, presenting the two intercepts μ_T and μ_S, and treatment effects and β and α are presented.

Solutions for Fixed Effects

Effect	distcb	treat	Estimate	Standard Error	DF
distcb	BINA		0.4953	0.2042	186
distcb	GAUS		-5.5340	1.2683	186
treat*distcb	BINA	1	-0.6566	0.2974	186
treat*distcb	GAUS	1	-2.8338	1.8743	186
treat*distcb	BINA	2	0	.	.
treat*distcb	GAUS	2	0	.	.

While in the bivariate vector of outcomes per subject the Gaussian outcome measured at six months, preceded the binary outcomes measured at one year, here the binary parameters preceed their Gaussian counterparts. This is merely because the levels in the 'distcb' variable are ordered alphabetically. So again, some care is needed.

Let us now switch to the random-effects models. Focusing on fitting model (24.19) to the ARMD data, the non-linear parameter λ prohibits the use of the GLIMMIX procedure, whence the procedure NLMIXED can be used. It is instructive to first focus on the case of two continuous outcome. In this case, the following three programs produce exactly the same model fit:

```
/*First program*/

proc mixed data=armd77 method=ml;
class treat time subject;
model visual = time treat*time
      / noint solution ddfm=satterthwaite;
random intercept
       / subject=subject type=un g v vcorr;
run;
```

```
/* Second program */

proc nlmixed data=armd77 qpoints=20 maxiter=50;
if time=1 then  eta = beta11 + b + beta12*(2-treat);
else if time=2 then eta = beta21 + b + beta22*(2-treat);
model visual ~ normal(eta,sigma*sigma);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
estimate 'sigma^2' sigma*sigma;
run;

/* Third program */

proc nlmixed data=armd77 qpoints=20 maxiter=50;
if time=1 then do;
   mean = beta11 + b + beta12*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          - 0.5*(visual-mean)**2/(sigma**2);
   ll = dens;
end;
else if time=2 then do;
   mean = beta21 + b + beta22*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          - 0.5*(visual-mean)**2/(sigma**2);
   ll = dens;
end;
model visual ~ general(ll);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
estimate 'sigma^2' sigma*sigma;
run;
```

Although the programs increase in terms of complexity and, for normally distributed outcomes, the first one perfectly does the job, they also increase the flexibility, but only the last one generalizes to joint outcomes. Indeed, the second one still is based on the assumption of a common outcome distribution, albeit with a differently defined mean structure. In the third one, the general likelihood feature is used and hence a different one can be used for each outcome separately.

Thus, a program of a continuous first outcome, combined with a binary second one, is as follows:

```
proc nlmixed data=armd77 qpoints=20 maxiter=100
             maxfunc=2000 technique=newrap;
parms beta11=-5.53 beta12=-2.83 beta21=-0.50
      beta22=0.66 sigma=7 lambda=3 tau=3;
```

```
if time=1 then do;
   mean = beta11 + lambda*b + beta12*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          -0.5*(contbin-mean)**2/(sigma**2);
   ll = dens;
end;
else if time=2 then do;
   eta = beta21 + b + beta22*(2-treat);
   p = exp(eta)/(1+exp(eta));
   ll = contbin*log(p) + (1-contbin)*log(1-p);
end;
model contbin ~ general(ll);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
estimate 'sigma^2' sigma*sigma;
run;
```

Reaching convergence is not straightforward, given the non-linear nature of the program, with the incorporation of λ, and a careful selection of starting values, and fine tuning using the convergence and updating method switches may be required.

Let us now turn attention to the MGLM and GLMM hierarchical models, presented in Section 24.4.3. The data need to be organized in a 'vertical' way, implying that the 4 continuous and 4 binary outcomes are stacked into a vector of length eight. An outprint for the first two patients:

Obs	subject	treat	repeat	time	dist	link	outcome
1	1	2	1	1	GAUS	IDEN	5
2	1	2	2	2	GAUS	IDEN	0
3	1	2	3	3	GAUS	IDEN	0
4	1	2	4	4	GAUS	IDEN	-10
5	1	2	5	1	BINA	LOGI	0
6	1	2	6	2	BINA	LOGI	1
7	1	2	7	3	BINA	LOGI	1
8	1	2	8	4	BINA	LOGI	1
9	2	1	1	1	GAUS	IDEN	-3
10	2	1	2	2	GAUS	IDEN	-3
11	2	1	3	3	GAUS	IDEN	-3
12	2	1	4	4	GAUS	IDEN	1
13	2	1	5	1	BINA	LOGI	1
14	2	1	6	2	BINA	LOGI	1
15	2	1	7	3	BINA	LOGI	1
16	2	1	8	4	BINA	LOGI	0

A program for the marginal model, using the GLIMMIX procedure, is

```
proc glimmix data=armd99 method=rspl empirical;
```

```
class time treat dist subject;
nloptions maxiter=50 technique=newrap;
model outcome = time*dist treat*time*dist
      / noint dist=byobs(dist) link=byobs(link) solution;
random _residual_ / subject=subject type=un r;
run;
```

which is a straightforward extension of the bivariate program.

Now, more work is needed to adapt the NLMIXED code for the conditional independence model:

```
proc nlmixed data=armd99 qpoints=20 maxiter=100
     maxfunc=2000 technique=newrap;
parms beta11=-1.55  beta12=1.00
      beta21=-2.93  beta22=1.02
  beta31=-6.68  beta32=1.52
  beta41=-13.47 beta42=2.81
  beta51=1.17  beta52=-0.22
  beta61=0.99  beta62=-0.47
  beta71=1.36  beta72=-0.41
  beta81=2.17  beta82=-0.11
  tau=1.77
  sigma=8.58
  lambda=-4.18
  ;
if repeat=1 then do;
   mean = beta11 + lambda*b + beta12*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          - 0.5*(outcome-mean)**2/(sigma**2);
   ll = dens;
end;
else if repeat=2 then do;
   mean = beta21 + lambda*b + beta22*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          - 0.5*(outcome-mean)**2/(sigma**2);
   ll = dens;
end;
else if repeat=3 then do;
   mean = beta31 + lambda*b + beta32*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
          - 0.5*(outcome-mean)**2/(sigma**2);
   ll = dens;
end;
else if repeat=4 then do;
   mean = beta41 + lambda*b + beta42*(2-treat);
   dens = -0.5*log(3.14159265358) - log(sigma)
```

```
          - 0.5*(outcome-mean)**2/(sigma**2);
   ll = dens;
end;
else if repeat=5 then do;
   eta = beta51 + b + beta52*(2-treat);
   p = exp(eta)/(1+exp(eta));
   ll = outcome*log(p) + (1-outcome)*log(1-p);
end;
else if repeat=6 then do;
   eta = beta61 + b + beta62*(2-treat);
   p = exp(eta)/(1+exp(eta));
   ll = outcome*log(p) + (1-outcome)*log(1-p);
end;
else if repeat=7 then do;
   eta = beta71 + b + beta72*(2-treat);
   p = exp(eta)/(1+exp(eta));
   ll = outcome*log(p) + (1-outcome)*log(1-p);
end;
else if repeat=8 then do;
   eta = beta81 + b + beta82*(2-treat);
   p = exp(eta)/(1+exp(eta));
   ll = outcome*log(p) + (1-outcome)*log(1-p);
end;
model outcome ~ general(ll);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
estimate 'sigma^2' sigma*sigma;
run;
```

Clearly, the code can be made a little more efficient in terms of programming code, but the advantage of the current program is clarity.

24.6 Concluding Remarks

We have discussed a number of methods to model correlated data when not all outcomes are of the same type. It is not uncommon to observe binary or otherwise categorical outcomes jointly with continuous outcomes, but also other combinations are perfectly possible. One might view such outcomes as multivariate. In addition, such a multivariate outcome of a heterogeneous nature can then be observed repeatedly over time, for various subjects within a trial, a cluster, or within other hierarchically organized units. Just as in the general case, we have distinguished between marginal, conditional, and random-effects models. A relatively large number of proposals have been made in the literature, many developed for specific applications.

We have focused on three modeling approaches in particular. The probit-normal and Plackett-Dale models are of a marginal nature and within the generalized linear mixed-effects modeling framework both marginal models, random-effects models, and random-effects models with residual or serial correlation can be considered. Each of these apply to a simple multivariate setting as well as to a fully hierarchical setting. In the literature, the marginal models have been combined with GEE and pseudo-likelihood ideas to enable parameter estimation when the outcome vectors are relatively long. In the GLMM framework, PQL and MQL can be used, as well as fully numerical integration. The examples have shown that these are feasible routes, but PQL and MQL are not recommended for random-effects models, due to the well-known bias issue. Therefore, numerical integration, such as in the SAS procedure NLMIXED, is a viable route. The SAS procedure GLIMMIX is useful for the purely marginal versions that can be seen as a version of GEE as well (Section 8.8).

In conclusion, thanks to recent software developments, the joint modeling of repeated measures of various outcome types can be done with standard statistical software and is not confined any more to user defined programming tools.