# Marginal correlation from logit- and probit-beta-normal models for hierarchical binary data



DOI: 10.1080/03610926.2012.709903
Handle: http://hdl.handle.net/1942/18574

# Marginal Correlation from Logit- and Probit-Beta-Normal Models for Hierarchical Binary Data

**Tony Vangeneugden** [ab] ** **Geert Molenberghs** [bc] **Geert Verbeke**[cb]

**Clarice G.B. Demétrio** [d]

[a] *Tibotec, Johnson & Johnson, Turnhoutseweg 30, B–2430 Beerse, Belgium*;

[b] *I-BioStat, Universiteit Hasselt, Agoralaan 1, B–3590 Diepenbeek, Belgium*;

[c] *I-BioStat, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium*;

[d] *ESALQ, Piracicaba, Saõ Paulo, Brazil*

**Abstract**

In hierarchical data settings, be it of a longitudinal, spatial, multi-level, clustered, or otherwise repeated nature, often the association between repeated measurements attracts at least part of the scientific interest. Quantifying the association frequently takes the form of a correlation function, including but not limited to intraclass correlation. Vangeneugden *et al* [1] derived approximate correlation functions for longitudinal sequences of general data type, Gaussian and non-Gaussian, based on generalized linear mixed-effects models (GLMM). Here, we consider the extended model family proposed by Molenberghs *et al* [2]. This family flexibly accommodates data hierarchies, intra-sequence correlation, and overdispersion. The family allows for closed-form means, variance functions, and correlation function, for a variety of outcome types and link functions. Unfortunately, for binary data with logit link, closed forms cannot be obtained. This is in contrast with the probit link, for which such closed forms can be derived. It is therefore that we concentrate on the probit case. It is of interest, not only in its own right, but also as a instrument to approximate the logit case, thanks to the well-known probit-logit 'conversion.' Next to the general situation, some important special cases such as exchangeable clustered outcomes receive attention because they produce insightful expressions. The closed-form expressions are contrasted with the generic approximate expressions of Vangeneugden *et al* [1] and with approximations derived for the so-called logistic-beta-normal combined model. A simulation study explores performance of the method proposed. Data from a schizophrenia trial are analyzed and correlation functions derived. **Some Keywords:** Beta-binomial model; Conjugate Random effects; Intraclass correlation; Generalized linear mixed model; Longitudinal data; Maximum likelihood; Probit link; Repeated measures.

## 1 Introduction

In applied sciences, one is often confronted with the collection of hierarchical data or repeated measures, in particular longitudinal, clustered and spatial data. For continuous data, the linear mixed model is commonly used (LMM, Verbeke and Molenberghs [3]). Marginalizing the linear mixed model over its random effects, and deriving marginal mean, variance, and correlation functions is straightforward. For example, psychometric

** Corresponding author.Email:tvangene@its.jnj.com

validation based on the LMM is rather feasible (Vangeneugden *et al* [4, 5]).

For non-Gaussian outcomes, this simplicity vanishes. First, one has to choose between marginal models, such as generalized estimating equations (GEE, Liang and Zeger [6]), or random-effects models, such as the generalized linear mixed model (GLMM, Breslow and Clayton [7]). It is generally found that marginalizing the latter is less than straightforward. Stumbling blocks are the non-linear nature of the link function and the mean-variance relationship (Molenberghs and Verbeke [8], Chapter 16). In spite of these issues, it is a viable modeling candidate, even when the marginal correlation is of interest. Vangeneugden *et al* [1]( henceforth abbreviated as VMLGBS) showed that the derivation of correlations based on GLMM is generally feasible if one is prepared to accept a Taylor-series based approximation. Their results are general, not only across data types, such as continuous outcomes, binary or ordinal outcomes, and counts, but also covers multivariate repeated measures, where more than one sequence per subject is measured repeatedly, even with different data types for the various sequences. In addition to their general results, these authors paid particular attention to binary outcomes, as well as on combined binary and Gaussian sequences.

One obvious though important case where the approximate results of VMLGBS become exact is for normally distributed outcomes, since then the GLMM reduces to the LMM. Molenberghs, Verbeke, and Demétrio [9](henceforth abbreviated as MVD) presented a model for longitudinal or otherwise hierarchical count data that simultaneously incorporates normal random effects in the linear predictor, as in any GLMM, as well as conventional overdispersion parameters. Overdispersion arises when the mean-variance relationship stemming from the posited generalized linear model (McCullagh and Nelder [10]) is too restrictive. This issue can occur with both cross-sectional as well as with hierarchical data. It is commonly encountered with count data. One convenient way to incorporate overdispersion is through gamma random effects, giving rise in the univariate case to the so-called negative binomial model (Breslow [11]). Hence, both random effects are combined into a single model. Thus, the proposal by MVD, referred to as the *combined* model, generalizes at once the GLMM and negative-binomial models and therefore, *a fortiori*, the univariate Poisson model. MVD also showed that, unlike the general GLMM case, the Poisson case allows for closed-form expressions for the mean vector, variance-covariance matrix, and even for the full joint probability vectors. This is true for the combined model and hence also for all of the aforementioned special cases, providing the opportunity to derive closed-form expressions for the within-unit correlation functions since it implies that there is no practical need for the Taylor-series based approximations mentioned earlier. Vangeneugden *et al* [12] have contrasted the generic approximate expressions for the correlation of repeated count data derived in Vangeneugden *et al* [1] with the results of the exact intra-sequence correlation expressions. They concluded that the proposed extension strongly outperforms the classical GLMM.

In their paper, Molenberghs *et al* [2](henceforth abbreviated as MVDV), proposed a broad class of generalized linear models accommodating overdispersion and clustering through two separate sets of random effects. They emphasize conjugate random effects at the level of the mean for the first aspect and normal random effects

embedded within the linear predictor for the second aspect, even though the proposed family is more general. They use the term *strong conjugacy* for these cases where conjugacy properties still apply when also normally distributed random effects are present. MVDV noted that, in comparison to the longitudinal Poisson case, the longitudinal binary case, with logit link, defeats closed-form solutions and strong conjugacy. In spite of it being a very natural choice in the univariate case, it does not combine very nicely with normal random effects. MVDV have shown that closed-form expressions can be derived for the probit case. Rodríguez and Elo [13] also noted the comfort of using the probit link, stemming from the existence of closed-form correlations, unlike with the logit and complementary log-log links. These authors mention that often the correlation on the (latent) scale of the linear predictor is calculated, rather than at the probability (manifest) scale. While the former is definitely easier to calculate, the latter may be the one of scientific interest. Moreover, there can be a large discrepancy between both correlations.

The paper is organized as follows. In Section 2, a motivating case study, stemming from clinical trials in patients with schizophrenia, is introduced. Section 3 contains four main ingredients: standard generalized linear models, with emphasis on the binary logistic and probit model; (2) classical models for overdispersion; (3) models for repeated measures with normal random effects, with focus on the binary-normal model; and (4) the general combined model of MVDV (Section 3.1) with further details for the Bernoulli case offered in Section 3.2. In Section 4, closed-form correlation functions are derived for the particularly insightful case of exchangeable clustered data. In Section 5, a Taylor-series-based approximation is derived. In Section 6, the results of the different models are contrasted and finally, Section 7 reports simulations, directed at evaluating the quality of the approximation and the closed formulae.

## 2 A Clinical Trials in Schizophrenic Patients

We have available individual patient data from four double-blind randomized clinical trials, comparing the effects of risperidone to conventional anti-psychotic agents for the treatment of chronic schizophrenia. Schizophrenia has long been recognized as a heterogeneous disorder with patients suffering from both "negative" and "positive" symptoms. Negative symptoms are characterized by deficits in social functions such as poverty of speech, apathy and emotional withdrawal. Positive symptoms entail more florid symptoms such as delusions and hallucinations, which are superimposed on the mental status. Several measures can be considered to assess a patient's global condition. The *Clinical Global Impression* (CGI) of overall change *versus* baseline is a 7-grade scale used by the treating physician to characterize how well a subject has improved since baseline. The levels are: 'very much improved,' 'much improved,' 'minimally improved,' 'no change,' 'minimally worse,' 'much worse,' 'very much worse.' Clinical response is often defined as a CGI score of 'very much improved or 'much improved.' Given that the label in most countries recommends doses ranging within 4–6 mg/day, we include in our analysis only patients who received either these doses of risperidone or an active control

(haloperidol, perphenazine, or zuclopenthixol). Depending on the trial, treatment was administered for a duration of 6–8 weeks. For example, in the international trials by Peuskens *et al* [14], Marder and Meibach [15], and Hoyberg *et al* [16] patients received treatment for 8 weeks, while in the study by Huttunen *et al* [17] patients were treated over a period of 6 weeks. The sample sizes were 453, 176, 74, and 71, respectively. Measurements were taken at weeks 1, 2, 4, 6, and 8. In Section 6, the pooled data from these trials will be analyzed. While a choice to pool can be perceived as controversial, here the trials are comprised of very similar patients.

## 3 Modeling Framework

Our model for binary outcomes is based upon the generalized linear model and two of its extensions, the fist one to accommodate overdispersion, the second one to account for data hierarchies, such as in longitudinal data. We briefly review these building blocks.

A random variable $Y$ follows an exponential family distribution if the density is of the form

$$f(y) \equiv f(y|\eta,\phi) \;=\; \exp\left\{\phi^{-1}[y\eta - \psi(\eta)] + c(y,\phi)\right\}, \tag{1}$$

for a specific set of unknown parameters $\eta$ and $\phi$, and for known functions $\psi(\cdot)$ and $c(\cdot,\cdot)$. Often, $\eta$ and $\phi$ are termed 'natural parameter' (or 'canonical parameter') and 'dispersion parameter,' respectively. It is well known that

$$\mathsf{E}(Y) \;=\; \mu = \psi'(\eta), \tag{2}$$

$$\mathsf{Var}(Y) \;=\; \sigma^2 = \phi\psi''(\eta), \tag{3}$$

implying a mean-variance relationship: $\sigma^2 = \phi\psi''[\psi^{'-1}(\mu)] = \phi v(\mu)$, with $v(\cdot)$ the so-called variance function. Of central interest here is the Bernoulli distribution with success probability $P(Y=1)=\pi$, the density can be written as

$$f(y) = \exp\left\{y\ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right\}, \tag{4}$$

which implies that the Bernoulli distribution belongs to the exponential family, with natural parameter $\theta$ equal to the logit, i.e., $\ln[\pi/(1-\pi)]$ of $\pi$, scale parameter $\phi=1$, with mean $\mu=\pi$ and variance function $v(\pi)=\pi(1-\pi)$. The natural link function is the logit link, leading to the classical logistic regression model with $\ln[\pi_i/(1-\pi_i)] = \eta_i = \boldsymbol{x_i}'\boldsymbol{\beta}$. An alternative to the logit link is the probit, where $\eta_i = \Phi^{-1}(\pi_i)$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. This model is slightly less standard because the probit model is not the natural link, as we will see in Section 3.2, it has appeal in the overdispersed and/or repeated contexts.

When the standard exponential-family models constrain the mean-variance relationship, so-called overdispersion is introduced. Early reviews are provided by Hinde and Demétrio [18, 19] provide general treatments of

overdispersion. The Poisson case received particular attention by Breslow [11] and Lawless [20]. A natural step is to allow the overdispersion parameter $\phi \neq 1$, so that (3) produces $\text{Var}(Y) = \phi v(\mu)$. A convenient route is through a two-stage approach. Generally, the two-stage approach is made up of considering a distribution for the outcome, given a random effect $f(y_i|\theta_i)$ which, combined with a model for the random effect, $f(\theta_i)$, produces the marginal model:

$$f(y_i) = \int f(y_i|\theta_i) f(\theta_i) d\theta_i. \tag{5}$$

When the data are hierarchical, let $Y_{ij}$ denote the $j$th outcome measured for cluster (subject) $i$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and $\boldsymbol{Y_i}$ the $n_i$-dimensional vector of all measurements available for cluster $i$. It is then common to include normal random effects in the linear predictor of the generalized linear model, giving rise to the family known as generalized linear mixed model (Engel and Keen [27], Breslow and Clayton [7], Wolfinger and O'Connell [28]). Assume that, in analogy with (1), conditionally upon $q$-dimensional random effects $\boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$, the outcomes $Y_{ij}$ are independent with densities of the form

$$f_i(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\}, \tag{6}$$

with

$$\eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) = \eta[E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi})] = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i} \tag{7}$$

for a known link function $\eta(\cdot)$, with $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ $p$-dimensional and $q$-dimensional vectors of known covariate values, with $\boldsymbol{\xi}$ a $p$-dimensional vector of unknown fixed regression coefficients, and with $\phi$ a scale (overdispersion) parameter. Finally, let $f(\boldsymbol{b_i}|D)$ be the density of the $N(\boldsymbol{0}, D)$ distribution for the random effects $\boldsymbol{b_i}$. A commonly encountered member of this family is the logit-normal model, but one can also consider a probit-normal model.

### 3.1 A Model Combining Overdispersion-related and Normal Random Effects

Bringing together concepts from the overdispersion models and the normal random effects model, MVDV specified a broad class of general linear models for repeated data accommodating overdispersion and clustering through two separate sets of random effects, producing the following general family:

$$f_i(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}) = \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\}, \tag{8}$$

with notation similar to what was used in (6), but now with conditional mean

$$E(Y_{ij}|\theta_{ij}, \boldsymbol{b_i}) = \mu^c_{ij} = \theta_{ij}\kappa_{ij}, \tag{9}$$

where the random variable $\theta_{ij} \sim \mathcal{G}_{ij}(\xi_{ij}, \sigma^2_{ij})$, $\kappa_{ij} = g(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i})$, $\xi_{ij}$ is the mean of $\theta_{ij}$ and $\sigma^2_{ij}$ is the corresponding variance. Finally, as before, $\boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$. Write $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$. Unlike in Section **??**, we now have two different notations, $\eta_{ij}$ and $\lambda_{ij}$, to refer to the linear predictor and/or the natural parameter. The reason is that $\lambda_{ij}$ encompasses the random variables $\theta_{ij}$, whereas $\eta_{ij}$ refers to the 'GLMM part' only.

It is convenient, but not strictly necessary, to assume that the two sets of random effects, $\boldsymbol{\theta}_i$ and $\boldsymbol{b_i}$, are independent of each other. Regarding the components $\theta_{ij}$ of $\boldsymbol{\theta}_i$, three useful special cases result from assuming that: (1) they are independent; (2) they are correlated, implying that the collection of univariate distributions $\mathcal{G}_{ij}(\xi_{ij}, \sigma^2_{ij})$ needs to be replaced with a multivariate one; and (3) they are equal to each other, useful in applications with exchangeable outcomes $Y_{ij}$. In this paper, we will operate under (1); the other cases can be studied similarly.

The relationship between mean and natural parameter now is

$$\lambda_{ij} = h(\mu^c_{ij}) = h(\theta_{ij}\kappa_{ij}). \tag{10}$$

We can still apply standard GLM ideas to derive the mean and variance, combined with iterated-expectation-based calculations. For the mean, it follows that $E(Y_{ij}) = E(\theta_{ij})E(\kappa_{ij}) = E[h^{-1}(\lambda_{ij})]$. MVDV have derived approximate formulae using Taylor expansion around $\boldsymbol{b_i} = \mathbf{0}$ for mean, variance and covariance function.

### 3.2 Bernoulli-type Models for Binary data

In the binary case the general model described in Section 3.1 becomes

$$\begin{aligned} Y_{ij} &\sim \text{Bernoulli}(\pi_{ij} = \theta_{ij}\kappa_{ij}), & (11)\\ \kappa_{ij} &= \frac{\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}\right)}{1 + \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}\right)}. & (12) \end{aligned}$$

Unlike in the Poisson case (see MVD), closed forms for neither the mean nor the variance follow in this case. When only overdispersion random effects are included, especially when they are assumed to follow a beta distribution, conjugacy applies. However, the combination of beta-distributed and normal distribution precludes strong conjugacy.

MVDV have derived general approximate expressions for the mean, variance and the covariance:

$$\begin{aligned} E(Y_{ij}) &= E(\theta_{ij})E(\kappa_{ij}), & (13)\\ \text{Var}(Y_{ij}) &= E(\theta_{ij})E(\kappa_{ij}) - E(\kappa_{ij})^2E(\theta_{ij})^2, & (14)\\ \text{Cov}(Y_{ij}, Y_{ik}) &= E(\kappa_{ij}.\kappa_{ik}) \cdot \text{Cov}(\theta_{ij}, \theta_{ik}) + E(\theta_{ij})E(\theta_{ik}) \cdot \text{Cov}(\kappa_{ij}, \kappa_{ik}). & (15) \end{aligned}$$

These authors further consider (13)–(15) for the three situations described in the previous section, i.e., the $\theta_{ij}$ generally correlated, uncorrelated, and equal. They also derived approximate expressions for (13)–(15) in case of $\theta_{ij} \sim \text{Beta}(\alpha, \beta)$ and contemplated into special cases.

In comparison to the longitudinal Poisson case, the just described longitudinal binary case defeats closed-form solutions and strong conjugacy. As stated before, more progress is possible with the probit link. The random-effects probit model has received some attention in earlier decades (Schall [30], Guilkey and Murphy [31],

Hedeker and Gibbons [32], McCulloch [33], Gibbons and Hedeker [34], Renard, Molenberghs, and Geys [35], Rodríguez and Elo [13]), with emphasis primarily on computational schemata to deal with the multivariate normal integral.

Replacing the normal with the probit, while at the same time assuming that the overdispersion parameters are beta distributed, comes down to:

$$\begin{aligned} \kappa_{ij} &= \Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}), \\ \theta_{ij} &\sim \text{Beta}(\alpha, \beta). \end{aligned} \tag{16}$$

Like before, $\alpha$ and $\beta$ could be allowed to vary with $i$ and/or $j$. MVDV have shown that the joint distribution can be written as:

$$f_{n_i}(\boldsymbol{y}_i = \mathbf{1}) = \left(\frac{\alpha}{\alpha+\beta}\right)^{n_i} \cdot \Phi_{n_i}(X_i\boldsymbol{\xi}; L_{n_i}^{-1}), \tag{17}$$

with

$$L_{n_i} = I_{n_i} - Z_i\left(D^{-1} + Z_i'Z_i\right)^{-1}Z_i'. \tag{18}$$

While this is the joint probability for a vector of nothing but successes, $(1,\ldots,1)'$, all other probabilities can be derived straightforwardly by appropriate contrasts of success probabilities:

$$f_{n_i}[\boldsymbol{y}_i = \boldsymbol{m}_i = (m_{i1},\ldots,m_{in_i})'] = \sum_{\boldsymbol{s}\supset\lambda(\boldsymbol{m}_i)} \text{sgn}(\boldsymbol{s})\Phi_{\#\boldsymbol{s}}\left(\widetilde{X}_i^{(\boldsymbol{s})}\boldsymbol{\xi}; L_{(\boldsymbol{s})}^{-1}\right) \cdot \left(\frac{\alpha}{\alpha+\beta}\right)^{\#\boldsymbol{s}}, \tag{19}$$

with $\lambda(\boldsymbol{m}_i) = \lambda(m_{i1},\ldots,m_{in_i})$ the set of places for which $m_{ij} = 1$,

$$\text{sgn}(\boldsymbol{s}) = \begin{cases} 1 & \text{if } \#\boldsymbol{s} - \#\lambda(\boldsymbol{m}_i) \text{ is even,} \\ 0 & \text{otherwise,} \end{cases}$$

$\widetilde{X}_i^{(\boldsymbol{s})}$ contains the rows from $X_i$ with row number in $\boldsymbol{s}$, and $L_{(\boldsymbol{s})}$ is the $\#s$-dimensional matrix built from the appropriate sub-matrices of these used in (18). The above developments evidently generalize when (16) is replaced with $\theta_{ij} \sim \text{Beta}(\alpha_j, \beta_j)$.

Next, the means, variances, and covariances can be derived from (17), by evaluating it for the one- and two-dimensional cases. We find:

$$E(Y_{ij}) = \frac{\alpha}{\alpha+\beta} \cdot \Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi}; L_1^{-1}) = \frac{\alpha}{\alpha+\beta} \cdot \Phi_1(|I + Dz_{ij}z'_{ij}|^{-1/2}\boldsymbol{x}'_{ij}\boldsymbol{\xi}), \tag{20}$$

$$\text{Var}(Y_{ij}) = \frac{\alpha}{\alpha+\beta} \cdot \Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi}; L_1^{-1}) \cdot \left[1 - \frac{\alpha}{\alpha+\beta}.\Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi}; L_1^{-1})\right], \tag{21}$$

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) = &\left(\frac{\alpha}{\alpha+\beta}\right)^2 \cdot \\ &\left\{\Phi_2\left[\begin{pmatrix} \boldsymbol{x}'_{ij} \\ \boldsymbol{x}'_{ik} \end{pmatrix}\boldsymbol{\xi}, L_{2jk}^{-1}\right] - \Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi}; L_{1j}^{-1})\Phi_1(\boldsymbol{x}'_{ik}\boldsymbol{\xi}; L_{1k}^{-1})\right\}, \end{aligned} \tag{22}$$

where

$$L_{2jk} = I_2 - \begin{pmatrix} \boldsymbol{z}'_{ij} \\ \boldsymbol{z}'_{ik} \end{pmatrix} \left[ D^{-1} + \begin{pmatrix} \boldsymbol{z}'_{ij} \\ \boldsymbol{z}'_{ik} \end{pmatrix} (\boldsymbol{z}_{ij}\ \boldsymbol{z}_{ik}) \right]^{-1} (\boldsymbol{z}_{ij}\ \boldsymbol{z}_{ik}),$$

Note that the rightmost density in (20) is standard normal. Evidently, (17) and (19) lead, not only to the mean, variance, and covariance expressions, but also to the higher-order moments.

## 4 Closed-form Derivation of the Correlation Function

In general, the correlation between two measurements $j$ and $k$ on the same experimental unit $i$ is:

$$\mathsf{Corr}(Y_{ij}, Y_{ik}) = \frac{\mathsf{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\mathsf{Var}(Y_{ij}) \cdot \mathsf{Var}(Y_{ik})}}. \tag{23}$$

For the binary case with probit link and assuming a Beta distribution for the overdispersion parameters, we can use (21) and (22) to derive a closed-form expression for the correlation. Because of its generality, it is hard to simplify (23), except in specific cases. We consider the special case of an exchangeable structure, where $\boldsymbol{x}'_{ij}\boldsymbol{\xi} = \xi$, and random intercepts, i.e. $\boldsymbol{z}_{ij} = 1$, and $D = d$. Then (21) and (22) simplify to

$$\mathsf{Var}(Y_{ij}) = \frac{\alpha}{\alpha+\beta} \cdot \Phi_1(\xi; L_1^{-1}) \cdot \left[ 1 - \frac{\alpha}{\alpha+\beta} . \Phi_1(\xi; L_1^{-1}) \right], \tag{24}$$

$$\mathsf{Cov}(Y_{ij}, Y_{ik}) = \left( \frac{\alpha}{\alpha+\beta} \right)^2 \cdot \left\{ \Phi_2 \left[ \begin{pmatrix} \xi \\ \xi \end{pmatrix}, L_{2jk}^{-1} \right] - [\Phi_1(\xi; L_{1j}^{-1})]^2 \right\}, \tag{25}$$

Now, via straightforward algebra it is easy to show that $L_{1j} = 1/(d+1)$ and

$$L_{2jk} = \frac{1}{1+2d} \begin{pmatrix} 1+d & -d \\ -d & 1+d \end{pmatrix}$$

Hence we have that

$$L_{2jk}^{-1} = \begin{pmatrix} 1+d & d \\ d & 1+d \end{pmatrix}. \tag{26}$$

Furthermore, we can transform to obtain the standard normal univariate and the bivariate distribution via the Cholesky decomposition:

$$\Phi_1(\xi; L_1^{-1}) = \Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right), \tag{27}$$

$$\Phi_2 \left[ \begin{pmatrix} \xi \\ \xi \end{pmatrix}, L_{2jk}^{-1} \right] = \Phi_2 \left[ \begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho \right], \tag{28}$$

where the correlation for the bivariate normal distribution based on (26) is defined by $\rho = d/(d+1)$.

Based on (24)–(25) and (27)–(28), the closed-form expression for the correlation becomes:

$$\text{Corr}(Y_{ij}, Y_{ij}) = \frac{\frac{\alpha}{\alpha+\beta} \cdot \left\{ \Phi_2 \left[ \begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho \right] - \left[ \Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \right]^2 \right\}}{\Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \cdot \left[ 1 - \frac{\alpha}{\alpha+\beta} \cdot \Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \right]}. \tag{29}$$

As a special case, consider the normal-probit model without Beta random effects, then (29) reduces to

$$\text{Corr}(Y_{ij}, Y_{ij}) = \frac{\Phi_2 \left[ \begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho \right] - \left[ \Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \right]^2}{\Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \cdot \left[ 1 - \Phi_1 \left( \frac{\xi}{\sqrt{d+1}} \right) \right]}. \tag{30}$$

## 5 Taylor-series-based Derivation of the Correlation Function

VMLGBS derived approximate expressions for the correlation function in the GLMM, including when multiple sequences on the same subject are observed. For convenience, these derivations are reproduced in Appendix A. Given the absence, for the entirely general case, of closed-form expressions for the moments and hence, *a fortiori*, for the joint distribution, this is pragmatically a sensible way forward. In our probit case, it is strictly speaking unnecessary to resort to such approximations. Their importance lies in the following: it is generally understood that Taylor-series-based expansions in GLMM could be highly inaccurate, especially when restricted to the first order. This has been amply documented for so-called penalized and marginal quasi-likelihood (for a summary, see Molenberghs and Verbeke [8]). However, most experience is built regarding marginal regression parameters in the logistic-linear case, whereas here we deal with correlations in the probit-linear case. This is why we offer such calculations. Moreover, because here it is possible to explicitly derive closed forms next to approximations, the agreement or discrepancy between them provides, with due caution, indirect evidence as to the quality of approximation in the logistic case.

When we consider the special case of an exchangeable structure as we did in Section 4, then $V_{ij}(0) = V(0) = \exp(\xi)/[1+\exp(\xi)]^2$ and (A.7) simplifies to

$$\widetilde{rho}_{ijk} = \text{Corr}(Y_{ij}, Y_{ij}) = \frac{V(0)d}{1+V(0)d}. \tag{31}$$

Similar approximate formulas can be derived for the combined logistic-Beta-normal model; we refer to Section A.2 of the Appendix for more detail. Upon using (A.8), (A.9), (A.10), and (23), with the same notation as above and writing $\kappa^{(0)} = \kappa$, then:

$$\widetilde{rho}_{ijk} = \frac{\left(\frac{\alpha}{\alpha+\beta}\right) \cdot \kappa^2 (1-\kappa)^2 d}{\left[\kappa + \kappa(1-\kappa)(1-2\kappa)\frac{1}{2}d\right] - \left(\frac{\alpha}{\alpha+\beta}\right) \cdot \left[\kappa + \kappa(1-\kappa)(1-2\kappa)\frac{1}{2}d\right]^2} \tag{32}$$

**Table 1:** *Estimated Intraclass Correlation Coefficient using probit and logistic models with and without over dispersion random effect.*

| Model | Dispersion | Estimate (s.e.) |
|---|---|---|
| Logistic-Normal (31) | No | 0.48 (0.026) |
| Logistic-Combined-Conjugate-Normal (32) | Beta | 0.62 (0.126) |
| Probit-Normal (30) | No | 0.67 (0.028) |
| Probit-Combined-Conjugate-Normal (29) | Beta | 0.60 (0.035) |

## 6 Analysis of the Schizophrenia Data

VMLGBS analyzed the data introduced in Section 2. They used the GLMM framework, applying a logistic model with normal random effects and have calculated that the intraclass coefficient of correlation was 0.48 (s.e. 0.026), using the approximation derived in (31). This figure needs to be approached, however, with due caution. To underscore this, we re-analyzed the data using the logistic model combining conjugate and normal random effects introduced in Section 3.1 with approximations provided in (32). Furthermore, we used the closed-form expressions (29) and (30), derived in Section 4. Table 1 summarizes results. First, the point estimate is questionable in the logistic case, due to the often poor approximation. This is clearest when there is no conjugate random effect. Furthermore, the standard error in the logistic case can be less than trustworthy as well; this phenomenon is now clearest in the combined case. It is therefore safer to use the probit-based models for correlation assessment, and other inferences. While both probit-based models are similar in terms of the correlation estimate, it also follows that ignoring the overdispersion aspect leads to a certain amount of overestimation for the correlation coefficient.

## 7 Simulation Study

To corroborate theoretical considerations and findings from the above data analysis, we performed a limited simulation study to evaluate the performance of the closed-form expressions from the probit model. We refer to VMLGBS for simulations to evaluate the approximate derivation for (31) using the Bahadur model for that purpose. We assume an exchangeable structure, where $\boldsymbol{x}_{ij}'\boldsymbol{\xi} = \xi$, and random intercepts, i.e., $\boldsymbol{z}_{ij} = 1$, and $D = d$. Admittedly, this is relatively limited, but it leads to a constant marginal correlation coefficient. Otherwise, the marginal correlation would be a function of the covariates.

**Table 2:** *Results of the simulation study.*

| True $\rho$ | $\alpha$ | $\xi$ | $d$ | $P_{00}$ | $P_{10} = P_{01}$ | $P_{11}$ | $N$ | Est. (s.d.) |
|---|---|---|---|---|---|---|---|---|
| 0.15 | 3 | -0.8 | 0.25 | 0.64 | 0.14 | 0.07 | 200 | 0.180 (0.146) |
| | | | | | | | 1000 | 0.177 (0.084) |
| 0.25 | 17 | -0.2 | 0.5 | 0.38 | 0.18 | 0.25 | 200 | 0.254 (0.111) |
| | | | | | | | 1000 | 0.257 (0.048) |
| 0.50 | 2.5 | -0.2 | 4.25 | 0.48 | 0.12 | 0.28 | 200 | 0.595 (0.074) |
| | | | | | | | 1000 | 0.600 (0.034) |
| 0.75 | 23.5 | -0.2 | 9 | 0.46 | 0.06 | 0.41 | 200 | 0.764 (0.063) |
| | | | | | | | 1000 | 0.777 (0.027) |

If we simulate 2 repeated measures, we can use (17) and (27) and (28) to derive that:

$$
\begin{aligned}
P_{11} &= \left(\frac{\alpha}{\alpha+\beta}\right)^2 \cdot \Phi_2\left[\begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho\right], \\
P_{01} = P_{10} &= \left(\frac{\alpha}{\alpha+\beta}\right) \cdot \Phi_1\left(\frac{\xi}{\sqrt{d+1}}\right) - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \cdot \Phi_2\left[\begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho\right], \\
P_{00} &= 1 - 2 \cdot \left(\frac{\alpha}{\alpha+\beta}\right) \cdot \Phi_1\left(\frac{\xi}{\sqrt{d+1}}\right) + \left(\frac{\alpha}{\alpha+\beta}\right)^2 \cdot \Phi_2\left[\begin{pmatrix} \frac{\xi}{\sqrt{(d+1)(1+2d)}} \\ \frac{\xi}{\sqrt{d+1}} \end{pmatrix}, \rho\right],
\end{aligned}
$$

where $\rho = d/d+1$ as before and where $P_{11}$, $P_{10}$, $P_{01}$, and $P_{11}$ indicate the four success/failure combinations. We used Monte Carlo machinery to simulate different true values for $\rho$: 0.15, 0.25, 0.50, and 0.75, respectively. For all 4 situations, 1000 datasets were generated, of size 200 and 1000, respectively. To obtain true $\rho$ values of 0.25, 0.50 and 0.75, specific values for $\alpha$, $d$, $\xi$ where determined and $\beta$ was fixed at $1/\alpha$. Table 2 summarizes the results.

Overall, the behavior of the estimator is quite acceptable, especially for sample sizes of 1000, where the average over the simulation run is invariably close to the true value. Note that there is no perfect match, because the covariates are chosen to approximate the targeted true value, without fully coinciding with it.

## 8 Concluding Remarks

Many inferential questions can be framed in terms of correlations between repeated or otherwise hierarchical measures taken on the same experimental unit. Such data can be conveniently modeled by means of random effects models, like linear and generalized linear mixed models. In contrast with the LMM case, the GLMM renders more difficult the derivation of such correlations. VMLGBS derived approximate correlation expressions in a broad GLMM framework, where not only a single one but several repeated-measures sequences per unit may be recorded, perhaps of a different data type. These authors paid explicit attention to binary endpoints, as well as to combinations of binary and Gaussian outcomes.

In this paper, we built on the work of MVDV who described a family of generalized linear models for repeated measures with normal and overdispersion random effects, the latter of conjugate. These authors have shown that in the case of binary data, the probit model allows to derive closed-form expressions for mean, variance and covariance. First, we derived approximate expressions based for the binary case, but with the probit link used rather than the more conventional logit link. For the overdispersion random effects, the beta distribution was chosen. Closed-form and approximate expressions were derived.

The quality of the approximations was assessed using data from a four schizophrenia trials. This is important to gauge whether the more conventional but computationally more involved logit-link-based models can be approximated by the convenient probit case. We concluded that the approximation is not all that good when only normal random effects are included, but that this radically changes when both sets of random effects are present; indeed, then the approximation's quality is excellent. Simulations based on the probit-beta-normal model showed that these exact derivations were stable and on average close to the true correlation.

Hence, the probit-beta-normal model provides a proper alternative to the logistic-beta-normal model, enabling exact derivations for the mean, variance, covariance and correlation due to strong conjugacy. Especially when correlation is of interest, this model should be contemplated.

## Acknowledgments

Financial support from the IAP research network #P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. The fourth author is supported by CNPq, a Brazilian Science Funding Agency.

## References

[1] T. Vangeneugden, G. Molenberghs, A. Laenen, H. Geys, C. Beunckens, and C. Sotto, *Marginal correlation in longitudinal binary data based on generalized linear mixed models*, Communications in Statistics, Theory & Methods (2009) **00**, 000–000.

[2] G. Molenberghs, G. Verbeke, C. Demétrio, and A. Vieira, *A Family of Generalized Linear Models for Repeated Measures With Normal and Conjugate Random Effects*, Submitted for publication (2009).

[3] G. Verbeke, G. and Molenberghs, *Linear Mixed Models for Longitudinal Data*, New York, Springer-Verlag, 2000.

[4] T. Vangeneugden, A. Laenen, H. Geys, D. Renard, and G. Molenberghs, *Applying linear mixed models to estimate reliability in clinical trial data with repeated measurements*, Controlled Clinical Trials (2004), **25**, 13–30.

[5] T. Vangeneugden, A. Laenen, H. Geys, D. Renard, and G. Molenberghs, *Applying concepts of generalizability theory on clinical trial data to investigate sources of variation and their impact on reliability*, Biometrics (2005), **61**, 295–304.

[6] K.Y. Liang, and S.L. Zeger, *Longitudinal data analysis using generalized linear models*, Biometrika (1986), **73**, 13–22.

[7] N.E. Breslow, and D.G. Clayton, *Approximate inference in generalized linear mixed models*, Journal of the American Statistical Association (1993), **88**, 9–25.

[8] G. Molenberghs, and G. Verbeke, *Models for Discrete Longitudinal Data*, New York: Springer, 2005

[9] G. Molenberghs, G. Verbeke, and C. Demétrio, *An extended random-effects approach to modeling repeated, overdispersed count data*, Lifetime Data Analysis (2007), **13**, 513–531.

[10] P. McCullagh, and J.A. Nelder, *Generalized Linear Models*, London: Chapman & Hall, 1989.

[11] N. Breslow, *Extra-Poisson variation in log-linear models*, Applied Statistics (1984), **33**, 38–44.

[12] T. Vangeneugden, G. Molenberghs, G. Verbeke, C. Demétrio, *Marginal correlation from an extended random-effects model for repeated and overdispersed counts*, Journal of Applied Statistics (2010), **00**, 000–000.

[13] G. Rodríguez and I. Elo, *Intra-class correlation in random-effects models for binary data*, The Stata Journal (2003) **3**, 32–46.

[14] J. Peuskens, and the Risperidone Study Group, *Risperidone in the treatment of chronic schizophrenic patients: a multinational, multicentre, double-blind, parallel-group study versus haloperidol*, British Journal of Psychiatry (1995) **166**, 712–726.

[15] S.R. Marder, R.C. and Meibach, *Risperidone in the treatment of schizophrenia*, American Journal of Psychiatry (1994) **151**, 825–835.

[16] O.J. Hoyberg, C. Fensbo, J. Remvig, O. Lingjaerde, M. Sloth-Nielsen, and I. Salvesen, *Risperidone versus perphenazine in the treatment of chronic schizophrenic patients with acute exacerbations*, Acta Psychiatrica Scandinavica (1993) **88**, 395–402.

[17] M.O. Huttunen, T. Piepponen, H. Rantanen, L. Larmo, R. Nyholm, and V. Raitasuo, *Risperidone versus zuclopenthixol in the treatment of acute schizophrenic episodes: a double-blind parallel-group trial*, Acta Psychiatrica Scandinavica (1995) **91**, 271–277.

[18] J. Hinde, and C.G.B. Demétrio, *Overdispersion: Models and estimation*, Computational Statistics and Data Analysis(1998) **27**, 151–170.

[19] J. Hinde, and C.G.B. Demétrio *Overdispersion: Models and Estimation*, São Paulo: XIII Sinape, 1998.

[20] J. Lawless, *Negative binomial and mixed Poisson regression*, The Canadian Journal of Statistics (1987), **15**, 209–225.

[21] J.G. Skellam, *A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials*, Journal of the Royal Statistical Society, Series B (1948) **10**, 257–261.

[22] R.R. Bahadur, *representation of the joint distribution of responses to* $n$ *dichotomous items* In: *Studies in Item Analysis and Prediction,*, H. Solomon (Ed.). Stanford Mathematical Studies in the Social Sciences VI. Stanford, CA: Stanford University Press, 1961.

[23] J.R. Dale, *Global cross-ratio models for bivariate, discrete, ordered responses*, Biometrics (1986) **42**, 721–727.

[24] G. Molenberghs, and E. Lesaffre, *Marginal modelling of correlated ordinal data using a multivariate Plackett distribution*, Journal of the American Statistical Association (1994) **89**, 633–644.

[25] P.F. Thall, and S.C. Vail, *Some covariance models for longitudinal count data with overdispersion*, Biometrics (1990) **46**, 657–671.

[26] C.D. Dean, *Estimating equations for mixed-Poisson models*, in: *Estimating Functions*, V.P. Godambe (Ed.). Oxford: Oxford University Press, 1991.

[27] B. Engel, and A. Keen, *A simple approach for the analysis of generalized linear mixed models*, Statistica Neerlandica (1994), **48**, 1–22.

[28] R. Wolfinger, and M. O'Connell, *Generalized linear mixed models: a pseudo-likelihood approach*, Journal of Statistical Computation and Simulation (1993) **48**, 233–243.

[29] L. Fahrmeir,and G. Tutz, *Multivariate statistical modeling based on Generalized Linear Models (2nd ed.)*, New York: Springer, 2001.

[30] R. Schall, *Estimation in generalized linear models with random effects*, Biometrika (1991), **78**, 719–729.

[31] D.K. Guilkey, and J.L. Murphy, *Estimation and testing in the random effects probit model*, Journal of Econometrics (1993), **59**, 301–317.

[32] D. Hedeker, and R.D. Gibbons, *A random-effects ordinal regression model for multilevel analysis*, Biometrics (1994), **51**, 933–944.

[33] C.E. McCulloch, *Maximum likelihood variance components estimation for binary data*, Journal of the American Statistical Association (1994), **89**, 330–335.

[34] R.D. Gibbons, and D. Hedeker, *Random effects probit and logistic regression models for three-level data*, Biometrics (1997), **53**, 1527–1537.

[35] D. Renard, G. Molenberghs, and H. Geys, *A pairwise likelihood approach to estimation in multilevel probit models*, Computational Statistics and Data Analysis (2004), **44**, 649–667.

# Appendix

## A Taylor-series-based Derivation of the Correlation Function

### A.1 Models with Normal Random effects

We can usefully write the general model as $\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\mu}_i$, the conditional mean, given the random effects, can be written as $\boldsymbol{\mu}_i = h(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i)$, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known design matrices, $\boldsymbol{\beta}$ are fixed-effect parameters, $\boldsymbol{b}_i$ are random effects, and $h$ is a known link function. Finally, $\boldsymbol{\varepsilon}_i$ is the residual error component. We will now derive a general formula for the variance-covariance matrix of $\boldsymbol{Y}_i$ without any restriction on the distribution of the outcome variable nor on the complexity of the model, e.g., allowing for serial correlation or not. This maximizes the similarity with the case of continuous, normally distributed outcomes. However, a key distinction is that in the linear case there is no mean-variance link, whereas here the residual variance will follow from the mean. The variance covariance matrix can be written as:

$$\boldsymbol{V}_i = \mathsf{Var}(\boldsymbol{Y}_i) = \mathsf{Var}(\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i) = \mathsf{Var}(\boldsymbol{\mu}_i) + \mathsf{Var}(\boldsymbol{\varepsilon}_i) + 2\mathsf{Cov}(\boldsymbol{\mu}_i, \boldsymbol{\varepsilon}_i). \tag{A.1}$$

It is easy to show that $\mathsf{Cov}(\boldsymbol{\mu}_i, \boldsymbol{\varepsilon}_i) = \mathsf{Cov}[\mathsf{E}(\boldsymbol{\mu}_i|\boldsymbol{b}_i), \mathsf{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{b}_i)] + \mathsf{E}[\mathsf{Cov}(\boldsymbol{\mu}_i, \boldsymbol{\varepsilon}_i|\boldsymbol{b}_i)] = 0$ since the first term is 0 and the second term equals $\mathsf{E}[\mathsf{E}(\boldsymbol{\mu}_i - \mathsf{E}(\boldsymbol{\mu}_i))(\boldsymbol{\varepsilon}_i)|\boldsymbol{b_i}] = 0$ as $\boldsymbol{\mu}_i$ is a constant when conditioning on $\boldsymbol{b}_i$. For the first term in (A.1) we have, using a first-order Taylor series expansion around $\boldsymbol{b}_i = \boldsymbol{0}$:

$$\begin{aligned} \mathsf{Var}(\boldsymbol{\mu}_i) &= \mathsf{Var}(\boldsymbol{\mu}_i(\boldsymbol{\eta_i})) = \mathsf{Var}(\boldsymbol{\mu}_i(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i)) && \text{(A.2)} \\ &\cong \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta_i}}\frac{\partial \boldsymbol{\eta_i}}{\partial \boldsymbol{b}_i}\,\Big|_{\boldsymbol{b}_i=0}\right)\boldsymbol{D}\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta_i}}\frac{\partial \boldsymbol{\eta_i}}{\partial \boldsymbol{b}_i}\,\Big|_{\boldsymbol{b}_i=0}\right)' \cong \boldsymbol{\Delta}_i\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Delta}_i', && \text{(A.3)} \end{aligned}$$

where $\boldsymbol{\Delta}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta_i}}\Big|_{\boldsymbol{b}_i=0}$. For the second term in (A.1), we have:

$$\mathsf{Var}(\boldsymbol{\varepsilon}_i) = \mathsf{Var}[E(\boldsymbol{\varepsilon}_i|\boldsymbol{b}_i)] + E[\mathsf{Var}(\boldsymbol{\varepsilon}_i|\boldsymbol{b}_i)] = E[\mathsf{Var}(\boldsymbol{\varepsilon}_i|\boldsymbol{b}_i)] = \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{\Sigma}_i\boldsymbol{\Phi}^{\frac{1}{2}}, \tag{A.4}$$

where $\boldsymbol{\Phi}$ is a diagonal matrix with the overdispersion parameters along the diagonal. In case there are no overdispersion parameters, $\boldsymbol{\Phi}$ is set equal to the identity matrix. Expand the variance function $\boldsymbol{\Sigma}_i$ so that

$$\mathsf{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{A}_i{}^{\frac{1}{2}}\boldsymbol{R}_i\boldsymbol{A}_i{}^{\frac{1}{2}}\boldsymbol{\Phi}^{\frac{1}{2}}, \tag{A.5}$$

where $\boldsymbol{R}_i$ is the correlation matrix and $\boldsymbol{A}_i$ is a diagonal matrix containing the variances following from the generalized linear model specification of $\boldsymbol{Y}_{ij}$ given the random effects $\boldsymbol{b}_i = \boldsymbol{0}$, i.e., with diagonal elements $v(\mu_{ij}|\boldsymbol{b}_i = \boldsymbol{0})$. Using (A.3) and (A.5), we have the following expression for the variance-covariance matrix (A.1):

$$\boldsymbol{V}_i \cong \boldsymbol{\Delta}_i\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Delta}_i' + \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{A}_i{}^{\frac{1}{2}}\boldsymbol{R}_i\boldsymbol{A}_i{}^{\frac{1}{2}}\boldsymbol{\Phi}^{\frac{1}{2}}. \tag{A.6}$$

When the canonical link is used, we have $\boldsymbol{A}_i = \boldsymbol{\Delta}_i$ and (A.6) can be written as: $\boldsymbol{V}_i \cong \boldsymbol{\Delta}_i\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Delta}_i' + \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{\Delta}_i^{\frac{1}{2}}\boldsymbol{R}_i\boldsymbol{\Delta}_i^{\frac{1}{2}}\boldsymbol{\Phi}^{\frac{1}{2}}$. If in addition, conditional independence (no serial correlation) is assumed, then (A.6)

simplifies to: $\boldsymbol{V}_i \cong \boldsymbol{\Delta}_i \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i' \boldsymbol{\Delta}_i' + \boldsymbol{\Phi}^{\frac{1}{2}} \boldsymbol{\Delta}_i \boldsymbol{\Phi}^{\frac{1}{2}}$. Further, if we reduce the random-effects part to a random-intercept model, i.e., $\boldsymbol{Z}_i = \mathbf{1}$ and $\boldsymbol{D} = d$, then (A.6) reduces to $\boldsymbol{V}_i \cong \boldsymbol{\Delta}_i (d\boldsymbol{J}) \boldsymbol{\Delta}_i' + \boldsymbol{\Phi}^{\frac{1}{2}} \boldsymbol{\Delta}_i \boldsymbol{\Phi}^{\frac{1}{2}}$.

When we have a random-intercept model for binomial data with a logit link and assuming no overdispersion, $\boldsymbol{V}_i$ reduces to $\boldsymbol{V}_i \cong \boldsymbol{\Delta}_i (d\boldsymbol{J}) \boldsymbol{\Delta}_i' + \boldsymbol{\Delta}_i = \boldsymbol{\Delta}_i (d\boldsymbol{J} + \boldsymbol{\Delta}_i^{-1}) \boldsymbol{\Delta}_i'$. Furthermore, $\boldsymbol{\Delta}_i$ is a diagonal matrix with $V_{ij}(0)$ as diagonal elements, where the variance function $V_{ij}(0) = \mu_{ij} \mid_{\boldsymbol{b}_i=0} (1 - \mu_{ij} \mid_{\boldsymbol{b}_i=0})$, and therefore $\boldsymbol{V}_i \cong \mathrm{diag}(V_{ij}(0))[d\boldsymbol{J} + \mathrm{diag}(V_{ij}(0))^{-1}]\mathrm{diag}(V_{ij}(0))$. In other words, the variance-covariance matrix for subject $i$ is specified by the matrix with elements: $v_{ijj} = V_{ij}(0)[1 + V_{ij}(0)d]$, $v_{ijk} = dV_{ij}(0)V_{ik}(0), \quad (j \neq k)$. Based on these, we can determine a first-order approximation of the marginal correlation between time point $j$ and $k$, which is the intraclass correlation coefficient of reliability:

$$\rho_{ijk} = \mathrm{Corr}(Y_{ij}, Y_{ik}) = \frac{V_{ij}(0)V_{ik}(0)d}{\sqrt{\{V_{ij}(0)[1 + V_{ij}(0)d]\}\{V_{ik}(0)[1 + V_{ik}(0)d]\}}}. \tag{A.7}$$

Note that, when $d = 0$, then $\rho_{ijk} = 0$, and when $d \to \infty$, then $\rho_{ijk} \to 1$. Thus, the full positive correlation range is attainable, quite unlike marginal models for correlated binary data, that experience restrictions on the correlation parameter space to certain degrees. For a discussion, see Molenberghs and Verbeke (2005). No negative correlations can occur, which is entirely in line with the model's hierarchical nature, i.e., where $d$ is and remains interpretable as a variance. The related discussion for the case of linear mixed models can be consulted in Verbeke and Molenberghs (2000).

### A.2 Combined Models with Normal and Beta Random Effects

In the binary case, the general model takes the form (11) and (12) where $\theta_{ij} \sim \mathrm{Beta}(\alpha, \beta)$ and $\boldsymbol{b_i} \sim N(\mathbf{0}, D)$. MVDV have derived general approximations for marginal model elements, using Taylor expansion of $\kappa_{ij}$ around $\boldsymbol{b_i} = \mathbf{0}$. In this case we have that $E(Y_{ij}) = E(\theta_{ij})E(\kappa_{ij})$ and $E(\theta_{ij}) = \alpha/(\alpha + \beta)$. Then, the approximation of the mean, variance, and covariance can be written as:

$$\begin{aligned}
E(Y_{ij}) &\approx \frac{\alpha}{\alpha + \beta} \cdot \left[ \kappa_{ij}^{(0)} + \frac{1}{2} \kappa_{ij}^{(0)} (1 - \kappa_{ij}^{(0)})(1 - 2\kappa_{ij}^{(0)}) \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij} \right], \\
\mathrm{Var}(Y_{ij}) &\approx E(Y_{ij})[1 - E(Y_{ij})], \\
\mathrm{Cov}(Y_{ij}, Y_{ik}) &\approx \left( \frac{\alpha}{\alpha + \beta} \right)^2 \cdot \kappa_{ij}^{(0)} (1 - \kappa_{ij}^{(0)}) \kappa_{ik}^{(0)} (1 - \kappa_{ik}^{(0)}) \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ik}.
\end{aligned}$$

From this we can derive the correlation between measurements $j$ and $k$ for the same experimental unit, as in (23).

In case of an exchangeable structure with a random intercept model, then $\kappa_{ij}^{(0)} = \kappa_{ik}^{(0)} = \kappa$, and hence the formulas above for mean, variance, and covariance simplify to

$$E(Y_{ij}) \approx \frac{\alpha}{\alpha + \beta} \cdot \left[ \kappa + \frac{1}{2} \kappa (1 - \kappa)(1 - 2\kappa) d \right], \tag{A.8}$$

$$\mathrm{Var}(Y_{ij}) \approx E(Y_{ij})[1 - E(Y_{ij})], \tag{A.9}$$

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) \approx \left( \frac{\alpha}{\alpha + \beta} \right)^2 \cdot [\kappa(1 - \kappa)]^2 d. \tag{A.10}$$