On the Connections Between Bridge Distributions, Marginalized Multilevel Models, and Generalized Linear Mixed Models

Geert Molenberghs[1,2], Michael G. Kenward[3], Geert Verbeke[1,2], Samuel Iddi[2] & Achmad Efendi[2]

[1] I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium

[2] I-BioStat, KU Leuven, Leuven, Belgium

[3] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

Correspondence: Geert Molenberghs, I-BioStat, Universiteit Hasselt, Martelarenlaan 42, Hasselt 3500, Belgium. E-mail: geert.molenberghs@uhasselt.be

Abstract

Generalized linear mixed models (GLMM) are commonly used to analyze hierarchical data. Unlike linear mixed models, they do not automatically provide parametric marginal regression functions, while such functions are needed for population-averaged inferences. This issue has received considerable attention and here three approaches to address it are reviewed, expanded, and compared: (1) the closed-form expressions of the marginal moments and distributions for a variety of GLMMs, derived by Molenberghs et al. (2010), as well as an extension that accommodates overdispersion; (2) the marginalized multilevel models of Heagerty (1999); (3) the bridge distribution of Wang and Louis (2003), a form for the random-effects distribution that allows the conditional and hierarchical mean to be described by the same link function. Our derivations are for the identity link function, the log link, and a collection of links for binary data. We highlight a number of useful connections: (a) it is shown that the bridge distribution for data with a mean on the unit interval is unique; (b) the three approaches are different for unit-interval data with the logit link, but are connected for the probit link; for the latter, there exist closed forms; (c) further results are derived for the bridge distribution in the case of unit-interval data and a Student's t link; (d) in contrast to the unit-interval case, it is shown how large classes of distributions act as bridge distributions when an identity or a logarithmic link is adopted; (e) for these links, the three approaches are either identical or closely connected; (f) it is underscored for a random-intercepts model and logarithmic link, that the data contain no information about the particular distribution for the random intercept, given that the same fit to the data can be ascribed to an entire class of random-intercept distribution; (g) the implications of the difference between the unit-interval case on the one hand and the identity and logarithmic cases on the other, regarding sensitivity to model assumptions, are discussed.

Keywords: Cauchy distribution, identity link, logit link, log link, marginal interpretation, mixed models, mixture distribution, probit link, random effects, random-effects distribution, t distribution

1. Introduction

The statistical modeler often makes use of random effects to take into account overdispersion, subject-level heterogeneity, or both. Classical examples of generalized linear models with overdispersion random effects (McCullagh & Nelder, 1989; Hinde & Demétrio, 1998a, 1998b) include the Poisson model for counts with gamma random effects, producing the negative-binomial model, and the Bernoulli model for binary data with beta random effects, leading to the beta-binomial model.

When data are hierarchically organized, it is common practice to consider mixed-effects models (Laird & Ware, 1982; Breslow & Clayton, 1993; Wolfinger & O'Connell, 1993; Engel & Keen, 1994; Goldstein, 2002; Verbeke & Molenberghs, 2000; Molenberghs & Verbeke, 2005). Molenberghs et al. (2010) formulated models for non-Gaussian overdispersed and hierarchical data with two sets of random effects simultaneously, which they termed the *combined model* (CM).

While the Gaussian linear mixed model (LMM) leads to tractable expressions for marginal and conditional

distributions, the same is not true for generalized linear mixed models (GLMM) and their extensions. As a consequence of this, a lot of research has been done to address or circumvent this complexity. Rather than adding yet new methodology, it is our aim to provide a coherent framework for the disparate yet relevant material available. As we view it, there are three major axes that can be brought together. The first one, direct termed *marginalization*, is integration of the hierarchically specified model over the random effects. The second axis, termed a *marginalized multilevel model* (MMM), was proposed by Heagerty (1999) and Heagerty and Zeger (2000), who simultaneously specify the marginal mean and the mean conditional upon the random effects. The third axis, using the so-called *bridge distribution*, was proposed by Wang and Louis (2003, 2004) for binary data with a regression function including fixed effects and a random intercept. They succeeded in maintaining the same link function for the marginal and hierarchical means.

By contrast, it is not our aim to add to the extensive discussion on the comparative utility of subject-specific and marginal (population-averaged) inferences. The former are used, for example, when the focus is on an individual patient's prediction in view of prophylactic action, whereas the latter are of interest, for example, when the effectiveness of public-health interventions or novel treatments are to be assessed. For a discussion on these topics, we refer to Diggle et al. (2002).

Within this framework, we consider seven topics. (a) We extend the work of Wang and Louis (2003, 2004) for binary data, by showing that the bridge distribution in this case is unique, owing to the connection between link and cumulative distribution functions. (b) It is shown how the three approaches differ for the binary case. (c) We derive bridge distributions for several further links and for vector rather than scalar random effects. (d) Moving beyond the binary case, it is shown that for the identity and logarithmic links, vast classes of bridge distributions exist. (e) We establish a relationship between the three operations mentioned above: (1) marginalizing a GLMM or a CM; (2) finding the connector function for a MMM or a COMMM; and (3) deriving the bridge distribution. In particular, for the log and identity links, appropriate choices exist to make the three specifications coincide in some cases or exhibit close connections in others. (f) For the identity and logarithmic cases, it is shown that the data may contain little or even no information about the parametric form of the random-effects distribution. (g) We briefly consider the implications of this last result in terms of the sensitivity of these models and their resulting inferences to unverifiable assumptions about unobservables.

In summary, the results established here clarify relationships, similarities, and differences between seemingly disparate approaches to formulating hierarchical models for non-Gaussian data. In so doing, we gain insight into the fundamental differences between data types for which the mean function has support over a finite interval on the one hand, and those with support on the half line (logarithmic case) and entire real line (identity case) on the other. Even though remarks are made regarding estimation, our results are not primarily computational in nature. Rather, they help the modeler choose an appropriate model formulation in view of the estimands about which inferences are to be made. For the same reason, small-sample and other practical issues are mostly beyond the remit of this paper.

The rest of the paper is organized as follows. Motivating data examples are introduced in Section 2. In Section 3, each of the three specifications is reviewed in as much detail as is needed for the remainder of the paper. In particular, the three integral equations that come with these specifications are juxtaposed and scrutinized. In subsequent sections, the important situations arising for binary and proportion data (Section 4), the identity link (Section 5), and the log link (Section 6) are studied in detail. Implications for the identification of the random-intercepts distribution are given in Section 7. In Section 8, some remarks are made regarding parameter estimation. An illustration is described in Section 9. Technical background, details regarding estimation, and example SAS code are provided in electronically available Supplementary Materials.

2. Data Examples

Two sets of data are introduced, with binary and count outcomes, respectively. This will permit us to highlight the differences between the binary case and cases where the log link is used. The data will be analyzed in Section 9.

2.1 Binary Data: A Clinical Trial in Onychomycosis

These data, previously analyzed in several publications, come from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments, A and B, for toenail dermatophyte onychomycosis (TDO), described in full detail by Debacker et al. (1996). TDO is a common toenail infection, which is difficult to treat and affects more than 2% of western populations (Roberts, 1992). Anti-fungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthily. The development of

new-generation compounds has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with treatment A or with treatment B. In total, 2×189 patients were randomized. Subjects were followed during 12 weeks of treatment and followed further, up to a total of 48 weeks. Measurements were taken at baseline, every month during treatment, and every 3 months from then on, leading to a maximum of 7 measurements per subject. The outcome of interest was the severity of the infection, coded as 0 (not severe) or 1 (severe). The question of interest was whether the percentage of severe infections decreased over time, and whether that evolution was different for the two treatment groups.

In all cases, the fixed-effects predictor will take the form:

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} = \beta_0 + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j, \tag{1}$$

where T_i is an indicator for treatment (1 for the experimental compound and 0 for standard treatment) and t_j is time in months at the jth visit.

2.2 Count Data: A Clinical Trial in Epileptic Patients

These data are from a randomized, double-blind, parallel group multi-center study for the comparison of placebo with a new anti-epileptic drug (AED), in combination with one or two other AED's (Faught et al., 1996). Patients were randomized after a 12-week stabilization period for the use of AED's, and during which the number of seizures were counted. After that run-in period, 45 patients were assigned to the placebo group, 44 to the new treatment. Patients were measured weekly and followed (double-blind) during 16 weeks; thereafter they entered a long-term open-extension study. Some patients were followed for up to 27 weeks. The outcome of interest is the number of epileptic seizures experienced during the latest week, i.e., since the last time the outcome was measured. The research question is whether or not the new treatment reduces the number of epileptic seizures.

Let the number of epileptic seizures that patient i experienced during follow-up period j be Y_{ij} and t_j the time point of the jth measurement. Further, $T_i = 1$ for the active group and $T_i = 0$ for placebo. All models make use of the following predictor:

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} = \left(\beta_{00} + \beta_{01}t_j\right) \cdot (1 - T_i) + \left(\beta_{10} + \beta_{11}t_j\right) \cdot T_i. \tag{2}$$

3. Three Related Specifications

3.1 Generalized Linear Mixed Models (GLMM) and Combined Models (CM)

We begin by introducing the combined model (CM). Suppose a longitudinal outcome Y_{ij} is measured for each independent subject $i = 1, \ldots, N$ at occasion $j = 1, 2, \ldots, n_i$. The response for the ith subject, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$ is assumed to follow an exponential family distribution. In a conventional GLMM, the outcome Y_{ij} is modeled conditionally on a normal random effect $\boldsymbol{b_i}$ that captures correlation between repeated measures and some overdispersion. Molenberghs et al. (2010) introduced a second set of random-effects θ_{ij} to represent overdispersion in a more flexible way. Hence, the density of an individual outcome Y_{ij}, conditional on both sets of random effects, takes the generic form:

$$f_{ij}\left(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \lambda ij, \phi\right) = \exp\{\phi^{-1}\left[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})\right] + c(y_{ij}, \phi)\}. \tag{3}$$

The conditional mean is modeled as the product:

$$E\left(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \theta_{ij}\right) = \mu_{ij}^c = \psi'(\lambda_{ij}) = \theta_{ij}\kappa_{ij}, \tag{4}$$

where the overdispersion random effect follows a distribution $\theta_{ij} \sim \Theta_{ij}\left(\upsilon_{ij}, \sigma_{ij}^2\right)$, with mean υ_{ij} and variance σ_{ij}^2. Given that θ_{ij} enters the mean function directly, it has to preserve the range of the mean. Molenberghs et al. argued that a conjugate distribution is a convenient choice for reasons of computation and interpretation. For example, in many cases the derivation of marginal moments and distributions then simplifies. The other mean component is written

$$g(\kappa_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i, \tag{5}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are $n_i \times p$ fixed-effects and $n_i \times q$ random-effects design matrices, respectively. Their jth rows are denoted by $\boldsymbol{x}'_{ij}$ and z'_{ij}, respectively. $g(\cdot)$ is a link function and $\boldsymbol{\beta}$ and $\boldsymbol{b}_i \sim N(\boldsymbol{0}, D)$ are fixed and random effects, respectively. The two sets of random effects θ_{ij} and $\boldsymbol{b}_i$ are conveniently assumed to be independent of each other, although this constraint can be relaxed. All parameters can be estimated, for example, using maximum likelihood (Molenberghs et al., 2010).

www.ccsenet.org/ijsp
International Journal of Statistics and Probability
Vol. 2, No. 4; 2013

Depending on the type of outcome under investigation, the distribution of θ_{ij} can be chosen appropriately, such as a beta distribution for binary data or a gamma distribution for count and time-to-event data. Molenberghs et al. (2010) considered various such examples, some of which will return in Sections 4–6. They paid particular attention to the case where the distribution for θ_{ij} is conjugate for exponential family (3). For the case where also normal random effects are also present, Molenberghs et al. (2010) defined strong conjugacy, which essentially means that a simple, closed-form marginalization over the strongly conjugate random effect, but conditional on the normal random effect, applies.

When the CM does not contain random effects θ_{ij}, a conventional GLMM (Breslow & Clayton, 1993; Wolfinger & O'Connell, 1993) arises. In reverse, when no normal random effects are present, standard overdispersion models follow, such as the beta-binomial model for binary data, the negative-binomial model for count data, etc. (Hinde & Demétrio 1998a, 1998b). Marginalization of the model's mean structure is discussed in Section 3.4.

3.2 Marginalized Multilevel Models (MMM) and Combined Marginalized Multilevel Models (COMMM)

The general marginalized multilevel model due to Heagerty (1999) can be written as:

$$
\begin{align}
g(\mu_{ij}^m) &= x_{ij}'\beta, & (6) \\
g(\mu_{ij}^c) &= \Delta_{ij} + z_{ij}'b_i, & (7) \\
b_i &\sim F_b(0, D), & (8) \\
Y_{ij}|b_i &\sim F_{Y^c}\left(\mu_{ij}^c, \upsilon\right).
\end{align}
$$

Here, $F_b(0, D)$ is a distribution with mean zero and variance-covariance matrix D, while $F_{Y^c}\left(\mu_{ij}^c, \upsilon\right)$ is a distribution parameterized by, say, a location and scale parameter, which could but do not have to be the mean and variance. The marginal mean $\mu_{ij}^m = E(Y_{ij})$ depends on an $n_i \times p$ matrix of p linear predictors X_i through a link function $g(\cdot)$. Further, the conditional mean $\mu_{ij}^c = E(Y_{ij}|b_i)$ relates to the random variable b_i with distribution (8) and the function Δ_{ij} connects the marginal and conditional means through the same link function; the latter aspect could be relaxed if desired (Griswold & Zeger, 2004). The outcome Y_{ij}, given the random effects, is assumed to follow an exponential family model like (3) but now obviously with conditional expectation modeled as in (7). The function Δ_{ij} is obtained from solving an integral equation that will be spelled out in Section 3.4. The model parameters β have a marginal interpretation, so the use of MMMs or COMMS is more appropriate when population-averaged inferences are to be made.

By analogy with the CM described in the previous section, Iddi and Molenberghs (2011) extended the MMM by the inclusion of overdispersion random effects, i.e., by writing the conditional mean as $\theta_{ij}\kappa_{ij}$, as in (4), but with parametric form (7) rather than (5). This slightly changes the integral equation leading to the connector Δ_{ij}, as will be shown in Section 3.4. Iddi and Molenberghs (2011) show that the resulting COMMM can be estimated easily from the data.

3.3 Bridge Distributions

Wang and Louis (2003) specified a model through the constraint that the marginal mean and conditional mean are specified by identical link functions, with predictors that are the same up to a multiplicative factor ϕ and an offset k:

$$
g(\mu_{ij}^m) = k + \phi x_{ij}'\beta, \qquad g(\mu_{ij}^c) = x_{ij}'\beta + z_{ij}'b_i \tag{9}
$$

Specification (9) is similar to (6)–(7), but now the distribution of b_i is unknown, rather than the connector Δ_{ij} in (7). From (9) it follows that β has a marginal as well as a subject-specific interpretation, up to the factor ϕ. Wang and Louis (2003) focused on the binary case, where $g(\cdot)$ is, for example, the logit, probit, or complementary log-log function. They restricted the random-effects structure to a random intercept only. Wang and Louis (2004) allowed for covariate-dependent random effects by modeling the variance of the random intercept. In sections to follow, we will see that in a number of settings, vector random effects also admit bridge constructions. This is true for the probit link for binary data (Section 4.3), the identity link (Section 5), and the log link (Section 6). Other link functions for the binary case are examined in Section 4. Note that here a combined-model version can be considered as well. This would follow from multiplying $g(\mu_{ij}^c)$ by θ_{ij}. The resulting marginal mean would get the extra factor $E(\theta_{ij})$. We do not consider this further.

3.4 Three Connected Integral Equations

As is clear from the discussion above, the three approaches are different marginalization operations but they have

much in common. The connections between the approaches can be important when considering appropriate modeling frameworks in the light of the type of inferences to be drawn. To bring this out more clearly, we now juxtapose them. In all three approaches, the fixed-effects predictor $x'_{ij}\beta$, random-effects predictor $z'_{ij}b_i$, and link function $g(\cdot)$.

Marginalization of GLMM or CM Integral: The random-effects density $f(b_i)$ is given, while the marginal mean is unknown. This leads to an explicit integral:

$$\mu_{ij}^m = \int g^{-1}(x'_{ij}\beta + z'_{ij}b_i)f(b_i)db_i. \tag{10}$$

In the case that the combined model is considered rather than the GLMM, the overdispersion random effect distribution is also given and (10) becomes:

$$\mu_{ij}^m = E(\theta_{ij}) \int g^{-1}(x'_{ij}\beta + z'_{ij}b_i)f(b_i)db_i. \tag{11}$$

Apart from special cases, μ_{ij}^m will not be a simple function of the model parameters. Thus, while marginalization is possible, sometimes even analytically, this does not imply that marginal effects of scientific interest (e.g., treatment effect) are always available in terms of simple parametric functions. This motivates the use of the following two approaches.

Marginalized Multilevel Model Integral Equation: The random-effects density $f(b_i)$ is given, with the connector function Δ_{ij} unknown and identified through the relationships:

$$\mu_{ij}^m = g^{-1}(x'_{ij}\beta) = \int g^{-1}(\Delta_{ij} + z'_{ij}b_i)f(b_i)db_i. \tag{12}$$

When a combined model is considered, the distribution of the overdispersion random effect is also assumed to be known and (12) changes to:

$$\mu_{ij}^m = g^{-1}(x'_{ij}\beta) = E(\theta_{ij}) \int g^{-1}(\Delta_{ij} + z'_{ij}b_i)f(b_i)db_i. \tag{13}$$

In contrast to (10) and (11), (12) and (13) are integral equations, which implicitly define the unknown quantity, here the connector function Δ_{ij}.

Bridge Integral Equation: The random-effects density $f(b_i)$ and the constants k and ϕ are unknown but identified through:

$$\mu_{ij}^m = g^{-1}(k + \phi x'_{ij}\beta) = \int g^{-1}(x'_{ij}\beta + z'_{ij}b_i)f(b_i)db_i. \tag{14}$$

As before, the identifying relationship is the solution to an integral equation.

In what follows, it will be shown that the three operations are different from each other for most link functions on the unit interval. The probit link is a notable exception in this respect. For a variety of other data types, either the linear link or the log link is routinely considered. In these cases, the three operations either coincide or are intimately connected. Furthermore, while for binary data there is a unique bridge distribution for a given link function, for continuous or non-negative data with linear and log links, the class of bridge distributions is very large.

4. The Binary Case

We will first review the specific theory for the binary case, as derived by Wang and Louis (2003), and present some further results and reflections. For ease of exposition, the focus will be on the unit interval, but transformation of the results to other finite intervals is straightforward.

4.1 Bridge Distribution Functions for the Binary Case

For the binary case, bridge integral Equation (14) becomes a map between cumulative distribution functions, because $g^{-1}(\cdot)$ is a monotonically increasing map from the real line onto the unit interval. To emphasize this, we will write $H(\cdot) \equiv g^{-1}(\cdot)$ for this case. Thus, (14) becomes (Wang & Louis, 2003):

$$H(k + \phi\eta) = \int H(b + \eta)f(b)db. \tag{15}$$

www.ccsenet.org/ijsp
International Journal of Statistics and Probability
Vol. 2, No. 4; 2013

Here, $\eta \equiv \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ for ease of notation, and the derivation is focused on a random intercept only. We will briefly sketch the arguments of Wang and Louis. Details can be found in the original paper. The authors took derivatives on both sides of (15) with respect to η, leading to

$$\phi h(k + \phi\eta) = \int h(b + \eta)f(b)db, \tag{16}$$

with $h = H'$, the first derivative. This is a convolution: $h * f_{-b}(\eta) = \phi h(k + \phi\eta)$. The subscript $-b$ refers to sign reversal. They then transformed this equation to the Fourier domain and applied properties of the Fourier transform to yield: $\mathcal{F}f_{-b}(\xi) = e^{ik\xi/\phi}\mathcal{F}h(\xi/\phi)\big/\mathcal{F}h(\xi)$, where

$$(\mathcal{F}h)(\xi) = \int e^{-i\xi x}h(x)dx, \tag{17}$$

the Fourier transform. Clearly, (17) maps the density corresponding to the inverse link function to its characteristic function, as also noted by Wang and Louis (2003). Applying the inverse Fourier transform yields the generic solution to (15):

$$f(b) = \frac{1}{2\pi}\int e^{i(k/\phi - b)\xi}\frac{(\mathcal{F}h)(\xi/\phi)}{(\mathcal{F}h)(\xi)}d\xi. \tag{18}$$

They also showed that, for symmetric $h(\cdot)$, $k = 0$, and that $0 \leq \phi \leq 1$.

In addition, note that the existence and uniqueness of the bridge density is guaranteed. Existence holds whenever $h(\cdot)$ is integrable and non-degenerate, which is satisfied for all conventional links, precisely stemming from the connection with density functions. Uniqueness follows by construction and from the uniqueness of the Fourier transform.

It is very important to realize that the above results apply strictly to the case where the (inverse) link function is in a 1–1 relationship with a density function. It is tempting to conclude that this holds for binary data only, but the same arguments apply to proportions. That said, binary data and proportions are the two data types for which the above results follow in a natural way. Thus, these results apply to a wide class of link functions for two particular data types and, by linear transformation, to all situations where the mean ranges over a finite interval. An example of this is when a correlation coefficient would be modeled to make it depend on covariates. Whereas the results in the binary case are very specific, they are general for the identity (Section 5) and log (Section 6) links.

4.2 Logit Link

The combination of the logit link for binary hierarchical data and normally distributed random effects is arguably the most commonly encountered mixed-model setting for non-Gaussian data. At the same time, it is the most problematic one in numerical terms.

When marginalizing the GLMM or CM, the integral is:

$$E(Y_{ij}) = \int_\theta \int_b \theta_{ij}\text{expit}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i)\varphi(\boldsymbol{b}_i|\boldsymbol{0}, D)f(\theta_{ij})d\boldsymbol{b}_i d\theta_{ij}. \tag{19}$$

Here, $\varphi(\boldsymbol{b}_i|\boldsymbol{0}, D)$ is the zero-mean normal density with variance-covariance matrix D. In what follows, the parameter arguments will be suppressed from notation. As has been well documented (Breslow & Clayton, 1993; Wolfinger & O'Connnell, 1993; Zeger, Liang, & Albert, 1988; Molenberghs & Verbeke, 2005), this integral has no closed-form solution. A consequence of this is that the MMM (or COMMM) integral equation

$$\text{expit}(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) = \int_\theta \int_b \theta_{ij}\text{expit}(\Delta_{ij} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i)\varphi(\boldsymbol{b}_i)f(\theta_{ij})d\boldsymbol{b}_i d\theta_{ij} \tag{20}$$

has no closed-form solution for the connector Δ_{ij}. In neither (19) nor (20) does θ_{ij} pose a problem, given that its integral is explicit and can be replaced by $E(\theta_{ij})$. In the more conventional GLMM situation where there are no overdispersion random effects, θ_{ij} is removed from (19) and (20).

Should it be possible to solve (19) and (20) analytically, perhaps up to linear transformations, then the marginal mean on the one hand and the connector on the other would take simple forms in $\boldsymbol{x}'_{ij}\boldsymbol{\beta}$, and hence the normal density would act as a bridge density, a contradiction given the uniqueness stated in Section 4.1 and the fact that

Wang and Louis (2003) derived a quite different form. In particular, these authors focused on a random intercept only and solved:

$$\text{expit}(k + \phi \mathbf{x}'_{ij}\boldsymbol{\beta}) = \int_b \text{expit}(\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i) f(b_i) db_i. \tag{21}$$

Their solution, derived through (18) and Fourier transform operations, reads $k = 0$ and

$$f(b_i) = \frac{1}{2\pi} \frac{\sin(\phi\pi)}{\cosh(\phi b_i) + \cos(\phi\pi)}, \tag{22}$$

with

$$\phi = \left(1 + \frac{3}{\pi^2}d\right)^{-1/2}, \tag{23}$$

and d the random-intercept variance. Wang and Louis (2003) studied the properties of (21).

Griswold and Zeger (2004) noted that, once the bridge distribution is given, it can be cast in MMM form: $\text{expit}(\mathbf{x}'_{ij}\boldsymbol{\beta}) = \int_b \text{expit}\left(\Delta_{ij} + b_i\right) dB_l(0, d)$, where $\Delta_{ij} = \phi \mathbf{x}'_{ij}\boldsymbol{\beta}$ and $B_l(0, d)$ the logit bridge distribution with mean 0 and variance d.

4.3 Probit Link

In spite of the close connection between logistic and probit regression for univariate binary data, the numerical aspects in the probit case are much simpler than in the logistic situation when data are hierarchically organized and random effects are present. It follows from several sources (Zeger, Liang, & Albert, 1988; Griswold & Zeger, 2004; Molenberghs et al., 2010) that:

$$E(Y_{ij}) = \Phi\left(\phi_{ij} \cdot \mathbf{x}'_{ij}\boldsymbol{\beta}\right) = \int_b \Phi(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) \varphi(\mathbf{b}_i) d\mathbf{b}_i, \tag{24}$$

with

$$\phi_{ij} = \left(1 + \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right)^{-1/2}. \tag{25}$$

Expression (24) is the marginal mean for the probit-normal GLMM. In addition, the connector follows immediately:

$$\Delta_{ij} = \phi_{ij}^{-1} \mathbf{x}'_{ij}\boldsymbol{\beta} = \sqrt{1 + \mathbf{z}'_{ij} D \mathbf{z}_{ij}} \cdot \mathbf{x}'_{ij}\boldsymbol{\beta}. \tag{26}$$

Finally, from (24) we see that the normal density is the bridge with $k = 0$ and ϕ_{ij} as in (25). Note that ϕ depends on subject i and occasion j. However, if the random-effects structure consists of an intercept only, then ϕ_{ij} in (25) reduces to

$$\phi = (1 + d)^{-1/2}. \tag{27}$$

The resemblance to (23) is immediately clear. Wang and Louis (2003) showed that the normal density is the bridge for the random-intercept probit model from straightforward application of (18).

In conclusion, in contrast to the logit case, all three operations lead to closed forms and can be said to coincide, up to perhaps a multiplicative factor of the form (25).

When the CM is used to extend the probit-normal GLMM, then (24) becomes $E(Y_{ij}) = E(\theta_{ij}) \cdot \Phi\left(\phi \cdot \mathbf{x}'_{ij}\boldsymbol{\beta}\right)$. Similarly, for the corresponding COMMM, connector (26) becomes

$$\Delta_{ij} = \sqrt{1 + \mathbf{z}'_{ij} D \mathbf{z}_{ij}} \cdot \Phi^{-1}\left[E(\theta_{ij})^{-1} \cdot \Phi(\mathbf{x}'_{ij}\boldsymbol{\beta})\right]. \tag{28}$$

Expression (28) does not generally simplify and therefore is considerably more complex than (26). This results from the fact that the probit, in contrast to the identity and logarithmic links in Sections 5 and 6, respectively, does not allow for easy absorption of constants. Note that the closed-form solutions for the probit link can also be exploited when using the logit link, given that one function, when appropriately scaled, can approximate the other (Demidenko, 2004, p. 338). This feature is alluded to for the combined model in Molenberghs et al. (2010).

For the MMM case, Griswold and Zeger (2004) constructed a connector integral equation that combines both probit and logit links. Details on this are given in the Supplementary Materials (Section A).

4.4 Other Link Functions

Apart from the logit and probit link functions, Wang and Louis (2003) studied the complementary log-log and Cauchy links. Some detail on the complementary log-log link is given in the Supplementary Materials (Section B.1). An overview of the derivations for the Cauchy links can be found there as well (Section B.2). For the latter case, when a Cauchy link $H(\eta) = 1/\pi\,(\pi/2 + \arctan\eta)$ is used, the bridge is again Cauchy, with parameter $c^2 = (\phi^{-1} - 1)^2$, leading to $\phi = (c + 1)^{-1}$. Note that while the random-effects distribution has no finite moments, the conditional and marginal outcome distributions are Bernoulli and hence well defined.

A natural accompaniment to the probit and Cauchy links, is Student's t distribution, which has not been studied before in this context. The density is:

$$h_\nu(\eta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{\eta^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Hurst (1995) shows that the characteristic function is

$$(\mathcal{F}h_\nu)(\xi) = \frac{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right)\cdot\left(\sqrt{\nu}|\xi|\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}, \tag{29}$$

where $K_\alpha(\xi)$ is the modified Bessel function of the second kind, with index α (Abramowitz & Stegun, 1964, p. 375). Using generic expression (18) and (29), together with the fact that symmetry implies $k = 0$, the bridge distribution for the t link can be shown to be:

$$f_\nu(b) = \frac{1}{2\pi\phi^{\nu/2}}\int e^{-ib\xi}\frac{K_{\nu/2}\left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right)}{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right)}d\xi. \tag{30}$$

The connection to the two limiting cases is immediately clear. If $\nu = 1$ then the t distribution reduces to the Cauchy distribution and the characteristic function becomes:

$$(\mathcal{F}h_{1/2})(\xi) = \frac{K_{1/2}(|\xi|)\cdot|\xi|^{1/2}}{\Gamma(1/2)2^{-1/2}}. \tag{31}$$

Plugging the expressions $K_{1/2}(x) = \sqrt{\pi/2}e^{-x}x^{-1/2}$ and $\Gamma(1/2) = \sqrt{\pi}$ into (31) immediately produces $(\mathcal{F}h_{1/2})(\xi) = \exp(-|\xi|)$, as mentioned for the Cauchy case, which led to a Cauchy bridge. Likewise, it can be derived (Hurst 1995) that (29) converges to $(\mathcal{F}h_\infty)(\xi) = \exp\left(-0.5\xi^2\right)$, the characteristic function of the normal distribution, a fact that also follows from straightforward application of (17). Therefore, we can conclude that (30) offers a continuum of bridge random-effects distributions, with Cauchy and normal distributions as limiting cases. More details on this case be found in the Supplementary Materials (Section B.3).

5. The Identity Link

The most prominent instance of the identity link is undoubtedly the linear mixed model (Verbeke & Molenberghs, 2000), where the conditional mean function is linear in the fixed and random effects, the latter normally distributed. The three operations of Section 3.4 are now almost trivial.

First, marginalizing the linear mixed model is:

$$E(Y_{ij}) = \int (x'_{ij}\beta + z'_{ij}b_i)\varphi(b_i)db_i = x'_{ij}\beta.$$

Second, the connector function of the MMM is found through:

$$x'_{ij}\beta = \int (\Delta_{ij} + z'_{ij}b_i)\varphi(b_i)db_i = \Delta_{ij},$$

i.e., $\Delta_{ij} = x'_{ij}\beta$.

Third, we turn to the bridge distribution. Note that, in contrast to (16), there now is no relationship with a density function, because the mean ranges over the entire real line. Formally, the bridge integral equation becomes:

$$k + \phi x'_{ij}\beta = \int (x'_{ij}\beta + z'_{ij}b_i)f(b_i)db_i. \tag{32}$$

Clearly, (32) is satisfied by any integrable density $f(\boldsymbol{b}_i)$, with $k_{ij} = \boldsymbol{z}'_{ij}E(\boldsymbol{b}_i)$ and $\phi = 1$. When $f(\boldsymbol{b}_i)$ has mean zero, then further $k = 0$; this is a more desirable result, because then k is constant. In particular, $f(\boldsymbol{b}_i) = \varphi(\boldsymbol{b}_i)$ is a solution. Comparing these three results, we can safely assert that the triple of operations can be made to coincide for linear links with zero-mean normally distributed random effects. This was not true for unit-interval links, except for the probit link and normally distributed random effects, where a close connection exists.

In sharp contrast with the uniqueness occurring in the binary case, the solution in the identity-link case is not unique. To see this, assume that $\boldsymbol{b}_i$ in (32) follows a distribution with finite mean and density $f(\boldsymbol{b}_i)$, then the equation reduces to:

$$k + \phi \boldsymbol{x}'_{ij}\boldsymbol{\beta} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}E(\boldsymbol{b}_i), \tag{33}$$

producing $\phi = 1$ and $k = \boldsymbol{z}'_{ij}E(\boldsymbol{b}_i)$. Even more generally, if the conditional predictor is changed to a generalized additive structure $E(Y_{ij}|\boldsymbol{b}_i) = \gamma_1(\boldsymbol{x}'_{ij}, \boldsymbol{\beta}) + \gamma_2(\boldsymbol{z}'_{ij}, \boldsymbol{b}_i)$, with obvious notation, and the function $\gamma_2(\cdot, \cdot)$ has finite expectation over the random effects, then we find $\phi = 1$ and $k = E[\gamma_2(\boldsymbol{z}'_{ij}, \boldsymbol{b}_i)]$. In conclusion, *every random-effects distribution* that satisfies mild regularity conditions is a bridge distribution for a model with identity link. The predictor can be linear or even generalized additive in fixed and random effects. Note that we did not specify the conditional outcome distribution, but restricted specification to the mean. A counterexample of a bridge that does not satisfy the regularity conditions is the Cauchy distribution. Owing to its lack of finite moments, the integral in (32) is not well defined.

One might argue that the identity link leads to an almost trivial situation, owing to linearity, and therefore the above results are not surprising. In the next section, it will be shown that the log link shares many, though not all, of these attractive properties.

6. The Ubiquitous Log Link

Switching to counts, it is natural to consider a Poisson GLM for the model specification conditional upon the random effects. Similarly, for time to event outcomes, one often assumes an exponential or Weibull GLM. This typically implies the use of a logarithmic link function. Marginalizing the corresponding Poisson-normal GLM is then also straightforward (Zeger, Liang, & Albert, 1988; Molenberghs, Verbeke, & Demétrio, 2007):

$$E(Y_{ij}) = \int e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i} \varphi(\boldsymbol{b}_i)d\boldsymbol{b}_i = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}}. \tag{34}$$

When the CM is considered instead, the conditional mean function becomes

$$E(Y_{ij}|\boldsymbol{b}_i, \theta_{ij}) = \theta_{ij}e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i}$$

and then, slightly modifying (34), the marginal mean takes the form

$$\begin{aligned}
E(Y_{ij}) &= \int_\theta \int_b \theta_{ij}e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i}\varphi(\boldsymbol{b}_i)f(\theta_{ij})d\boldsymbol{b}_i d\theta_{ij} \\
&= E(\theta_{ij})e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}} = e^{\ln E(\theta_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}}.
\end{aligned} \tag{35}$$

Note that we are able to absorb the overdispersion random effect into the argument to the inverse link function, a consequence of strong conjugacy (Molenberghs et al., 2010). In the specific case that θ_{ij} follows a Gamma$(\alpha_{1j}, \alpha_{2j})$ distribution, the first term in (35) is $\ln(\alpha_{1j}\alpha_{2j})$.

Turning to the second operation, the connector function follows from (Griswold & Zeger, 2004):

$$e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}} = \int e^{\Delta_{ij} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i}\varphi(\boldsymbol{b}_i)d\boldsymbol{b}_i = e^{\Delta_{ij} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}}, \tag{36}$$

and hence

$$\Delta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}. \tag{37}$$

For the corresponding COMMM, embedding θ_{ij} into (36) produces the connector function

$$\Delta_{ij} = \ln E(\theta_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}. \tag{38}$$

For the bridge distribution, it is immediately clear that the bridge integral equation

$$e^{k_{ij}+\phi \boldsymbol{x}'_{ij}\boldsymbol{\beta}} = \int e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+z'_{ij}\boldsymbol{b}_i} f(\boldsymbol{b}_i)d\boldsymbol{b}_i \tag{39}$$

is satisfied for normally distributed random effects, $\phi = 1$, and $k_{ij} = \frac{1}{2}z'_{ij}Dz_{ij}$. It is even possible to incorporate the overdispersion random effect θ_{ij} into the model and then absorb the logarithm of its mean into the predictor and hence into k_{ij}.

As in the case of the identity link, the three operations are strongly related, though not identical. Indeed, for the marginalized GLMM, the marginalized CM, and for the bridge, the constant k_{ij} appears in the marginal mean, whereas the same constant appears in the connector when considering the MMM or COMMM.

Also in line with the identity link, a wide variety of random-effects distributions satisfy the bridge integral equation (39). Slightly generalizing this requirement:

$$e^{k_{ij}+\phi \gamma_1(\boldsymbol{x}_{ij},\boldsymbol{\beta})} = \int e^{\gamma_1(\boldsymbol{x}_{ij},\boldsymbol{\beta})+\gamma_2(\boldsymbol{z}_{ij},\boldsymbol{b}_i)} f(\boldsymbol{b}_i)d\boldsymbol{b}_i, \tag{40}$$

it is clear that *every* random-effects distribution satisfies (40), with $\phi = 1$ and

$$k_{ij} = \ln\left[\int e^{\gamma_2(\boldsymbol{z}_{ij},\boldsymbol{b}_i)} f(\boldsymbol{b}_i)d\boldsymbol{b}_i\right] = \ln E\left[e^{\gamma_2(\boldsymbol{z}_{ij},\boldsymbol{b}_i)}\right],$$

provided the expectation exists.

In the case of the identity link, the result follows from additivity of the integration operator. Here, a similarly general result follows from the fact that integration and multiplication with a constant can be interchanged; the fixed-effects portion of the mean acts as the multiplier.

At the beginning of this section, we referred to count data, but nowhere in the development was this feature used explicitly. Only the use of the log link is essential. This implies that the results are valid for data other than counts, for example they also apply to the Weibull, the gamma, and the inverse Gaussian models for continuous data, to name but a few. Some detail on these is given in the Supplementary Materials (Section C).

The relative ease with which marginal and conditional means have the same or similar parametric forms for the log link, as well as for the identity link, has been noted also by Ma and Jørgensen (2007), who propose the Tweedie exponential dispersion distribution to build generalized mixed models. For the log-link case, their family encompasses, for example, the Poisson, gamma, and inverse-Gaussian model. It also encompasses the normal model for the identity link. This implies, for example, that the bridge approach and the Tweedie approach in the Poisson case coincide.

7. Random-Intercepts Distributions in Models with Log Link

In this section, we use the log link and a random-intercepts model to provide a simple illustration of the fact that only the first moment of the random-effects distribution is identifiable from the data. Assume that the random intercept follows a distribution with density $f(b_i)$, then it follows that:

$$E(Y_{ij}) = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\ln E_f\left(e^{b_i}\right)}. \tag{41}$$

Further if the random intercept were normally distributed, (41) becomes:

$$E(Y_{ij}) = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\frac{1}{2}d}. \tag{42}$$

Clearly, the two random-intercepts related constants on the right hand sides of (41) and (42) must be identical and hence

$$E_f\left(e^{b_i}\right) = e^{\frac{1}{2}d}. \tag{43}$$

If $f(b_i)$ is a density with known parametric form but unknown parameters, then the left hand side of (43) can be equated to the right hand side with d replaced by $\widehat{d}$. In conclusion, the fit from a GLMM can be changed into that of a collection of models with alternative random-intercept distributions, (a) without re-fitting a model to the data and (b) without altering the marginal mean function. This result is connected to that of Verbeke and Molenberghs

(2011), who showed that the posterior distribution of the random effects given the data in hierarchical models is unverifiable from the data. The difference between our setting and theirs is that they considered the fit of the entire (multivariate) marginal model, whereas here we are concerned with the marginal mean function, an important component of the marginal model that nevertheless does not uniquely identify it. With our particular choice for the log link and a random intercept only, one can go further and show that, in the same way, the prior is unverifiable, because $\phi = 1$ and the random intercepts merely contribute a constant to the marginal linear predictor; this constant can stem from a variety of distributions, not just from a standard normal.

8. Estimation

All models considered can be fitted easily with available software, for example, the SAS procedure NLMIXED. In the Supplementary Materials (Section D), various approaches are reviewed and example code is given.

Every data type and link function has its own subtleties. It is clear from Section 4.2 that the three operations will lead to different models, in spite of the fact that each one can be implemented in standard software. By contrast, because the probit link corresponds to a normal bridge distribution, the models resulting from the three approaches (GLMM, MMM, bridge) all allow for closed forms. Nevertheless, the models themselves are not entirely equivalent. In a conventional GLMM, it is clear from (24) that the regression function $\widehat{\phi}_{ij} \boldsymbol{x}'_{ij} \widehat{\boldsymbol{\beta}}$ changes with levels of $\boldsymbol{z}_{ij}$, owing to the presence of ϕ_{ij}. In contrast, in a MMM, the connector function is introduced to cancel out such changes and hence the marginal regression function simply becomes $\boldsymbol{x}'_{ij} \widehat{\boldsymbol{\beta}}$. In other words, the choice between GLMM and MMM is between which of the two regression functions, either marginal or conditional, will depend on the random-effects model. Note that this is not an issue when only a random intercept is present, because then ϕ is constant. Thus, only for a random-intercept model are the three approaches exactly the same. For general random-effects structures, the GLMM and bridge models are identical, while these two are closely related but not identical to the MMM.

The log link setting is almost a mirror image to the probit link. Whereas for the probit link $k = 0$ and $\phi \neq 1$, showing that the GLMM and MMM are of the same functional form but with different parameterization, for the log link $k \neq 0$ and $\phi = 1$, as shown in Section 6. One can then fit a classical GLMM, for example a Poisson-normal model for count data. Evidently, the GLMM and the bridge are identical because the normal is one of the many possible bridge distributions. Furthermore, in contrast to the probit case, the GLMM and the MMM have the same marginal covariate effects, except for the intercept, as is clear from (37). The MMM absorbs the changes in intercept into the connector function. However, whenever the intercept is of no scientific interest, it is sufficient to fit a GLMM. This fact has been known for a long time and was reported in Zeger, Liang and Albert (1988). Moreover, the same holds between the CM and COMMM when also incorporating overdispersion random effects, as can be seen from (38). Given the many situations where a log link applies (see Section 6), this useful result is very broadly applicable.

9. Illustration

9.1 Binary Data: A Clinical Trial in Onychomycosis

Eight models are considered: standard GLMM with logit (1a) and probit (1b) link, as well as (1c) its CM extension with logit link; MMM with logit (2a) and probit (2b) link, and (2c) its COMMM extension with logit link; and finally bridge with logit (3a) and probit (3b) link. Molenberghs et al. (2010) also reported (1a) and (1c); Iddi and Molenberghs (2011) added (2a) and (2c); the remainder are added here. The code for (3a), using the SAS procedure NLMIXED, follows that of Wang and Louis (2003). For (2a), the marginal logit link is combined with a probit link for the conditional mean, for ease of implementation, as is clear from Section 4.3. Griswold and Zeger (2004) stated that the model where both links are of a logit type remains computationally challenging.

The fixed-effects predictor (1) is combined with a random intercept with $b_i \sim N(0, d)$. The overdispersion random effect is assumed to follow a Beta(α_1, α_2) distribution, in agreement with earlier analyses (Molenberghs et al., 2010; Iddi & Molenberghs, 2011).

Parameter estimates and standard errors are displayed in Table 1. Comparing panels (a) and (b) clearly shows the impact of changing the logit to the probit link. In panel (a), all three models are different, in agreement with our derivations in Section 4.2. In contrast, when a probit link is chosen, the GLMM and bridge models are identical, while the parameters in the MMM all follow from the same multiplicative transformation, by a factor $\sqrt{1 + d^2}$, a fact known from Section 4.3. The combined model extensions, displayed in panel (c) are given for generality. Comparing panels (b) and (c) shows that correcting for overdispersion reduces the strength of the treatment effect.

We have not considered a combined version of the bridge models for the binary case, because the mean of the overdispersion random effect cannot be absorbed into the predictor within the inverse link function, and hence it is not even clear how to define the bridge integral equation in an unambiguous fashion. This is different for the log link case, as shown in Section 6.

Table 1. A clinical trial in onychomycosis. Parameter estimates and standard errors. For the parameter of interest, β_3, also the p-value is given

Effect	Par.	Parameter estimates and standard errors		
(a) Models without overdispersion random effects (logit link)				
		(1a) GLMM	(2a) MMM	(3a) bridge
Intercept	β_0	-1.6306 (0.4345)	-0.6154 (0.1493)	-1.5103 (0.4126)
Treatment effect	β_1	-0.1146 (0.5852)	-0.0382 (0.2120)	-0.0856 (0.5694)
Time effect	β_2	-0.4041 (0.0459)	-0.1529 (0.0190)	-0.4139 (0.0475)
Interaction effect	β_3	-0.1613 (0.0718)	-0.0702 (0.0288)	-0.1687 (0.0760)
p-value		*0.025*	*0.016*	*0.027*
St. dev. R.I.	$\sqrt{d}$	4.0150 (0.3812)	2.1061 (0.1904)	0.2286 (0.0219)
(b) Models without overdispersion random effects (probit link)				
		(1b) GLMM	(2b) MMM	(3b) bridge
Intercept	β_0	-0.9193 (0.2290)	-0.3922 (0.0924)	-0.9193 (0.2290)
Treatment effect	β_1	-0.0786 (0.3075)	-0.0336 (0.1311)	-0.0786 (0.3075)
Time effect	β_2	-0.1961 (0.0212)	-0.0837 (0.0097)	-0.1961 (0.0212)
Interaction effect	β_3	-0.0773 (0.0326)	-0.0330 (0.0140)	-0.0773 (0.0326)
p-value		*0.018*	*0.018*	*0.018*
St. dev. R.I.	$\sqrt{d}$	2.1199 (0.1947)	2.1199 (0.1947)	2.1199 (0.1947)
(c) Models with overdispersion random effects (probit link)				
		(1c) CM	(2c) COMMM	—
Intercept	β_0	-1.3214 (1.5639)	-0.4762 (0.0408)	
Treatment effect	β_1	-1.3429 (2.1432)	-0.1858 (0.1240)	
Time effect	β_2	-1.6522 (0.4210)	-0.1832 (0.0241)	
Interaction effect	β_3	-0.7235 (0.3403)	-0.0691 (0.0392)	
p-value		*0.034*	*0.079*	
Std. dev. R.I.	$\sqrt{d}$	15.7371 (3.5433)	8.8901 (0.0152)	
Overdispersion	α_2/α_1	0.2828 (0.0372)	0.2769 (0.0363)	

9.2 Count Data: A Clinical Trial in Epileptic Patients

All six models fitted make use of the log link and are based upon the linear predictor (2). These are: (1a) the standard GLMM, (2a) the MMM, and (3a) the bridge model; (1b)–(3b) are the combined-model extensions. For the combined models, the overdispersion random effects are assumed to follow a Gamma(α, $1/\alpha$) distribution.

Parameter estimates are presented in Table 2. As shown in Section 6, the bridge and GLMM models coincide, because the normal distribution is one of the many possible bridges and the GLMM happens to be based, therefore, on one particular member of the bridge family. In the binary case though, the logistic-normal GLMM is not based on a bridge. This allows us to also consider a combined-bridge model (3b), in contrast to the binary case. Obviously, models (1a) and (3a) on the one hand, and (1b) and (3b) on the other, are identical. Furthermore, as shown in the same section, the marginalized versions of these coincide with the original model except for the intercept. This implies that all treatment-effect parameters and functions of these are independent of the choice between GLMM, MMM, or bridge (with the same holding true for the combined versions). Note that, while both intercepts change under marginalization, they do so to the same degree: $0.5 * d$, as follows from (37), but applied in the random-intercepts case. This implies that the difference between the two intercepts, interpretable as the treatment difference at baseline, does *not* change when marginalizing. In conclusion, for the log link, only the intercepts transform and do so in an additive way, whereas for the probit link the transformation is multiplicative and applies to all fixed-effects parameters.

Table 2. A clinical trial in patients with epileptic seizures. Parameter estimates and standard errors

Effect	Par.	Par. estimates and standard errors	
(a) Models without overdispersion random effects			
		(1a) GLMM &	
		(3a) bridge	(2a) MMM
Intercept placebo	β_{00}	0.8179 (0.1677)	1.3960 (0.1887)
Slope placebo	β_{01}	-0.0143 (0.0044)	-0.0143 (0.0044)
Intercept treatment	β_{10}	0.6475 (0.1701)	1.2256 (0.1901)
Slope treatment	β_{11}	-0.0120 (0.0043)	-0.0120 (0.0043)
Std. dev. R.I.	$\sqrt{d}$	1.0755 (0.0857)	1.0755 (0.0857)
(c) Models with overdispersion random effects			
		(1b) CM &	
		(3b) c-bridge	(2b) COMMM
Intercept placebo	β_{00}	0.9112 (0.1755)	1.4757 (0.1962)
Slope placebo	β_{01}	-0.0248 (0.0077)	-0.0248 (0.0077)
Intercept treatment	β_{10}	0.6555 (0.1782)	1.2200 (0.1970)
Slope treatment	β_{11}	-0.0118 (0.0075)	-0.0118 (0.0075)
Std. dev. R.I.	$\sqrt{d}$	1.0625 (0.0871)	1.0625 (0.0871)
Overdispersion	α	2.4640 (0.2113)	2.4640 (0.2113)

10. Concluding Remarks

In this paper, we have brought together three strands of research: (1) directly marginalizing the GLMM or its CM extension; (2) the MMM and the COMMM extension; and (3) the bridge distribution. The following observations wer made.

First, the special status of data on the unit interval, i.e., binary data and proportions, was demonstrated. In this case, the bridge distribution is unique, owing to the fact that a distribution function is used as inverse link function.

Second, within this class, the logit link has a special status, because the GLMM, MMM, and bridge approaches are fundamentally different from each other, and the first two do not allow for a closed-form solution.

Third, in contrast to this, the probit link does allow for closed-form solutions in all cases, given that the normal distribution is the bridge distribution. The bridge and GLMM approaches are identical, and closely connected to the MMM.

Fourth, additional link functions for the unit-interval case were studied, including the Cauchy and complementary log-log link, already studied by Wang and Louis (2003), and Student's t link, studied here for the first time. An appealing feature of this link is that the Cauchy link follows as a special case and the probit link follows by taking a limit.

Fifth, fundamental differences between the unit-interval case and all cases where a log link is used have been establishes. With this link, many random-effects distribution can be used as bridge distribution. Furthermore, the three approaches are identical, apart from a difference in effect on the intercept. It follows from this that the researcher has to reflect on which approach to choose, in view of the inferences to be drawn. For purely subject-specific inferences, the GLMM is a sensible choice, whereas the MMM is an obvious candidate for marginal inferences. The bridge models might be preferred when both marginal and conditional inferences are of interest. For the log and identity link cases, the choice is less critical, given the close connection between the model families.

Sixth, for the log link and in particular for the random-intercept model, it is shown that the marginal regression function contains no information about a particular random-intercept distribution, provided the latter has a finite mean.

We conclude by a note on generality. The results for link functions on the unit interval can be generalized, using appropriate linear transformations, to any finite interval. Likewise, the results for the logarithmic link can be made to apply to any half line, not just the non-negative real numbers. It was shown through the gamma and inverse Gaussian models that seemingly different links, such as the inverse and the squared inverse, turn into a version of

the logarithmic link when applying the range-preserving logarithmic transformation to the linear predictor. The generality of the identity link is self-evident.

Acknowledgements

References

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of Mathemaical Functions*. Washington, DC: National Bureau of Standards.

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

De Backer, M., De Keyser, P., De Vroey, C., & Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day - a double blind comparative trial. *British Journal of Dermatology, 134*, 16-17. http://dx.doi.org/10.1111/j.1365-2133.1996.tb15653.x

Demidenko, E. (2004). *Mixed Models*. New York: John Wiley & Sons. http://dx.doi.org/10.1002/0471728438

Diggle, P. J., Heagerty, P. J., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data* (2nd ed.). Oxford Science Publications. Oxford: Clarendon Press.

Engel, B., & Keen, A. (1994). A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica, 48*, 1-22. http://dx.doi.org/10.1111/j.1467-9574.1994.tb01428.x

Faught, E., Wilder, B. J., Ramsay, R. E., Reife, R. A, Kramer, L. D., Pledger, G. W., & Karim, R. M. (1996). Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosage. *Neurology, 46*, 1684-1690. http://dx.doi.org/10.1212/WNL.46.6.1684

Goldstein, H. (2002). *Multilevel Statistical Models* (3rd ed.). Oxford: Oxford University Press.

Griswold, M. E., & Zeger, S. L. (2004). *On marginalized multilevel models & their computation* (November 2004). Johns Hopkins University, Department of Biostatistics Working Paper #99.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics, 55*, 688-698. http://dx.doi.org/10.1111/j.0006-341X.1999.00688.x

Heagerty, P. J., & Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science, 15*, 1-26. http://dx.doi.org/10.1214/ss/1009212670

Hinde, J., & Demétrio, C. G. B. (1998a). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis, 27*, 151-170. http://dx.doi.org/10.1016/S0167-9473(98)00007-3

Hinde, J., & Demétrio, C. G. B. (1998b). *Overdispersion: Models and Estimation*. São Paulo: XIII Sinape.

Hurst, S. (1995). The characteristic function of the student *t* distribution. *Financial Mathematics Research Report No. FMRR 006–95, Statistics Research Report No. SRR044–95, Center for Financial Mathematics, School of Mathematical Sciences, Australian National University, Canberra, Australia.*

Iddi, S., & Molenberghs, G. (2011). A combined overdispersed and marginalized multilevel model. *Computational Statistics and Data Analysis, 56*, 1944-1951.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963-974. http://dx.doi.org/10.2307/2529876

Ma, R., & Jørgensen, B. (2007). Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B, 69*, 625-641. http://dx.doi.org/10.1111/j.1467-9868.2007.00603.x

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall/CRC.

Molenberghs, G., & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G., Verbeke, G., & Demétrio, C. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis, 13*, 513-531. http://dx.doi.org/10.1007/s10985-007-9064-y

Molenberghs, G., Verbeke, G., Demétrio, C., & Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science, 25*, 325-347. http://dx.doi.org/10.1214/10-STS328

Roberts, D. T. (1992). Prevalence of dermatophyte onychomycosis in the United Kingdom: Results of an omnibus survey. *British Journal of Dermatology, 126 Suppl. 39*, 23-27. http://dx.doi.org/10.1111/j.1365-2133.1992.tb00005.x

Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Verbeke, G., & Molenberghs, G. (2011). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling, 10*, 391-419. http://dx.doi.org/10.1177/1471082X0901000403

Wang, Z., & Louis, T. A. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika, 90*, 765-775. http://dx.doi.org/10.1093/biomet/90.4.765

Wang, Z., & Louis, T. A. (2004). Marginalizing binary mixed-effect models with covariate-dependent random effects and likelihood inference. *Biometrics, 60*, 884-891. http://dx.doi.org/10.1111/j.0006-341X.2004.00243.x

Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation, 48*, 233-243. http://dx.doi.org/10.1080/00949659308811554

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics, 44*, 1049-1060. http://dx.doi.org/10.2307/2531734

Supplementary Materials

A. Combining Probit and Logit Links into a Single Model

For the MMM case, Griswold and Zeger (2004) constructed a connector integral equation that combines both probit and logit links. Also allowing for overdispersion random effects as in the COMMM (Iddi & Molenberghs, 2011), the integral equation becomes:

$$\text{expit}(x'_{ij}\beta) = \int_\theta \int_b \theta_{ij}\Phi(\Delta_{ij} + z'_{ij}b_i)\varphi(b_i)f(\theta_{ij})db_i d\theta_{ij} = E(\theta_{ij}) \cdot \Phi(\phi_{ij}\Delta_{ij}), \tag{1}$$

leading to

$$\Delta_{ij} = \sqrt{1 + z'_{ij}Dz_{ij}} \cdot \Phi^{-1}\left[E(\theta_{ij})^{-1} \cdot \text{expit}(x'_{ij}\beta)\right]. \tag{2}$$

When there are no overdispersion random effects, (1) simplifies to

$$\Delta_{ij} = \sqrt{1 + z'_{ij}Dz_{ij}} \cdot \Phi^{-1}\left[\text{expit}(x'_{ij}\beta)\right]. \tag{3}$$

We note the following three points.

First, this MMM operation does not correspond to marginalization of a GLMM or CM, because integrating the right hand side of (3) with prespecified $\Delta_{ij} = x'_{ij}\beta$ produces a probit, which would lead us back to (24). This makes it genuinely a different operation, useful in its own right.

Second, this operation does not correspond to an easy bridge construction, as can be shown by contradiction as follows. Assuming that the normal density would act as a bridge, we would have a solution for k and ϕ in

$$\text{expit}(k + \phi x'_{ij}\beta) = \Phi\left(\frac{x'_{ij}\beta}{\sqrt{1 + z'_{ij}Dz_{ij}}}\right) = \Phi\left(x'_{ij}\widetilde{\beta}\right),$$

where the notation $\widetilde{\beta}$ is self-evident. While $k = 0$ follows immediately, the solution for ϕ would be

$$\phi = (x'_{ij}\beta)^{-1}\text{logit}\left[\Phi\left(x'_{ij}\widetilde{\beta}\right)\right],$$

which depends on β and hence is not allowable. Of course, as a partial answer to this, it is possible to extend the general bridge distribution theory by replacing (15) with a version where H on the left and right hand side is different, H_1 and H_2 say. This would lead to replacing (18) by

$$f(b) = \frac{1}{2\pi} \int e^{i(k/\phi-b)\xi} \frac{(\mathcal{F}h_1)(\xi/\phi)}{(\mathcal{F}h_2)(\xi)} d\xi.$$

Further, moving to links not on the unit interval, the generic bridge integral Equation (14) can be replaced by

$$\mu_{ij}^m = g_1^{-1}(k + \phi x'_{ij}\beta) = \int g_2^{-1}(x'_{ij}\beta + z'_{ij}b_i)f(b_i)db_i.$$

The fact that there is no corresponding bridge construction implies that this MMM operation stands on its own.

Third, maintaining the probit link on the right hand side of (3) but replacing the expit on the left hand side by a generic c.d.f. $H(\cdot)$, yields an explicit connector, in the spirit of (2):

$$\Delta_{ij} = \sqrt{1 + z'_{ij}Dz_{ij}} \cdot \Phi^{-1}\left[E(\theta_{ij})^{-1} \cdot H(x'_{ij}\beta)\right].$$

Thus, the closed form is entirely due to the use of the probit link in the conditional model, and is independent of the link for the marginal model.

B. Details on Other Link Functions for the Binary Case

B.1 The Complementary Log-Log Link

For the complementary log-log $H(\eta) = 1 - \exp[-\exp(\eta)]$, Wang and Louis (2003) derived the bridge:

$$f(b) = \frac{1}{2\pi} \int e^{i(k/\phi - b)\xi} \frac{\Gamma\left(1 - \frac{i\xi}{\phi}\right)}{\Gamma(1 - i\xi)} d\xi,$$

with $\phi = \left(1 + 6\pi^{-2}d\right)^{-1}$. In other words, the log-positive stable distribution is the bridge for the complementary log-log link. $\Gamma(\cdot)$ is the gamma function.

B.2 The Cauchy Link

When the Cauchy distribution

$$H(\eta) = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan \eta\right)$$

is used, the corresponding density is $h(\eta) = \left[\pi(1 + \eta^2)\right]^{-1}$, with Fourier transform $(\mathcal{F}h)(\xi) = \exp(-|\xi|)$, a well-known special Fourier transform (Spiegel, 1968, p. 176, Section 33.18). The bridge density then simply follows:

$$\begin{aligned}
f(b) &= \frac{1}{2\pi}\int_0^{+\infty} e^{-ib\xi - \frac{\xi}{\phi} + \xi}d\xi + \frac{1}{2\pi}\int_{-\infty}^0 e^{-ib\xi + \frac{\xi}{\phi} - \xi}d\xi \\
&= \frac{1}{2\pi}\left(\frac{-1}{1 - \phi^{-1} - ib} - \frac{1}{1 - \phi^{-1} + ib}\right) \\
&= \frac{1}{\pi}\frac{(\phi^{-1} - 1)}{(\phi^{-1} - 1)^2 + b^2}.
\end{aligned}$$

Hence, the standard Cauchy link produces a Cauchy bridge with parameter $c^2 = (\phi^{-1} - 1)^2$, leading to $\phi = (c+1)^{-1}$. Note that while the random-effects distribution has no finite moments, the conditional and marginal outcome distributions are Bernoulli and hence well defined.

B.3 The t Link

We can also consider Student's t distribution as an inverse link function, with corresponding density

$$h_\nu(\eta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{\eta^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Hurst (1995) shows that the characteristic function is

$$(\mathcal{F}h_\nu)(\xi) = \frac{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right) \cdot \left(\sqrt{\nu}|\xi|\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}, \tag{4}$$

where $K_\alpha(\xi)$ is the modified Bessel function of the second kind, with index α (Abramowitz & Stegun, 1964, p. 375). which implies that

$$(\mathcal{F}h_\nu)\left(\frac{\xi}{\phi}\right) = \frac{K_{\nu/2}\left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right) \cdot \left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}, \tag{5}$$

The ratio of (4) and (5) is

$$\frac{(\mathcal{F}h_\nu)\left(\frac{\xi}{\phi}\right)}{(\mathcal{F}h_\nu)(\xi)} = \frac{K_{\nu/2}\left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right)}{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right) \cdot \phi^{\nu/2}} \tag{6}$$

Using generic expression (18) and (6), together with the fact that symmetry implies $k = 0$, the bridge distribution for the t link can be shown to be:

$$f_\nu(b) = \frac{1}{2\pi\phi^{\nu/2}} \int e^{-ib\xi} \frac{K_{\nu/2}\left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right)}{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right)} d\xi. \tag{7}$$

The link with the two limiting cases is immediately clear. If $\nu = 1$ then the t distribution reduces to the Cauchy distribution and the characteristic function becomes:

$$(\mathcal{F}h_{1/2})(\xi) = \frac{K_{1/2}(|\xi|) \cdot |\xi|^{1/2}}{\Gamma(1/2)2^{-1/2}}. \tag{8}$$

Plugging the expressions $K_{1/2}(x) = \sqrt{\pi/2}e^{-x}x^{-1/2}$ and $\Gamma(1/2) = \sqrt{\pi}$ into (8) immediately produces $(\mathcal{F}h_{1/2})(\xi) = \exp(-|\xi|)$, as mentioned for the Cauchy case, which led to a Cauchy bridge. Likewise, it can be derived (Hurst, 1995) that (4) converges to $(\mathcal{F}h_\infty)(\xi) = \exp(-0.5\xi^2)$, the characteristic function of the normal distribution, a fact that also follows from straightforward application of (17).

Therefore, we can conclude that (7) offers a continuum of bridge random-effects distributions, with Cauchy and normal distributions as limiting cases.

C. Further Uses of the Log Link

When the log link is applied to the eclectic collection of models made up of the Poisson, Weibull, gamma, inverse Gaussian model, etc. then a large class of distributions acts as bridge distribution on the one hand, and the three operations can be made to coincide for normal random effects on the other, as was stated in Section 6. Given that the Poisson model was studied in detail in the main paper, we provide some further material on the Weibull, gamma, and inverse Gaussian cases.

In the Weibull case, the conditional mean of the CM can be written as (Molenberghs et al., 2010):

$$E(Y_{ij}|\boldsymbol{b}_i, \theta_{ij}) = \Gamma\left(1 + \frac{1}{\rho}\right)\theta_{ij}\lambda^{-1/\rho}\exp\left(-\frac{1}{\rho}\boldsymbol{x}'_{ij}\boldsymbol{\beta} - \frac{1}{\rho}\boldsymbol{z}'_{ij}\boldsymbol{b}_i\right). \tag{9}$$

Here, ρ is the Weibull shape parameter, and λ is an additional mean parameter, that could be absorbed in either the linear predictor or the overdispersion random effect θ_{ij}. Given normality of $\boldsymbol{b}_i$, the marginal mean follows easily:

$$E(Y_{ij}) = \Gamma\left(1 + \frac{1}{\rho}\right)E(\theta_{ij})\lambda^{-1/\rho}\exp\left(-\frac{1}{\rho}\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2\rho^2}\boldsymbol{z}'_{ij}D\boldsymbol{z}'_{ij}\right). \tag{10}$$

From this marginalization, and by analogy with the Poisson derivations above, the connector is:

$$\Delta_{ij} = -\frac{1}{\rho}\boldsymbol{x}'_{ij}\boldsymbol{\beta} - \frac{1}{2\rho^2}\boldsymbol{z}'_{ij}D\boldsymbol{z}'_{ij} - \ln E(\theta_{ij}). \tag{11}$$

The last term on the right hand side of (11) applies only when the COMMM is considered. Likewise, also here the normal distribution is one of the many bridge distributions. If adopted, it corresponds to the values $\phi = 1$ and $k = \boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}/(2\rho^2)' + \ln E(\theta_{ij})$.

When a gamma model is adopted (McCullagh & Nelder, 1989; Krishnamoorty, 2006, Ch. 15; Evans, Hastings, & Peacock, 2000, Ch. 19), it can be written in exponential family form as:

$$f(y_{ij}) = \exp\left[-\alpha_{1ij}\ln\alpha_{2ij} - \ln\Gamma(\alpha_{1ij}) - \frac{1}{\alpha_{2ij}}y_{ij} + (\alpha_{1ij} - 1)\ln y_{ij}\right].$$

The natural parameter can be written as

$$\theta_{ij} = -\alpha_{2ij}^{-1} = -\exp\left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i\right], \tag{12}$$

based upon which the mean becomes

$$\mu_{ij} = \alpha_{1ij}\alpha_{2ij} = -\frac{\alpha_{1ij}}{\theta_{ij}} = e^{\ln\alpha_{1ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{b}_i}, \tag{13}$$

and hence, like in the Weibull case, the log link results apply.

Finally, the same machinery can be applied to the Inverse Gaussian distribution (McCullagh & Nelder, 1989; Krishnamoorty, 2006, Ch. 27; Evans, Hastings, & Peacock, 2000, Ch. 22), of which the exponential-family form is

$$f(y_{ij}) = \exp\left[-\lambda\frac{y_{ij}}{2\mu_{ij}^2} + \frac{\lambda}{\mu_{ij}} + c(y_{ij}, \lambda)\right] = \exp\left[-\lambda\left(\theta_{ij}y_{ij} - \sqrt{2\theta_{ij}}\right) + c(y_{ij}, \lambda)\right], \tag{14}$$

with

$$c(y_{ij}, \lambda) = -\frac{\lambda}{2y_{ij}} + \frac{1}{2}\ln\lambda - \frac{1}{2}\ln(2\pi) - \frac{3}{2}\ln y_{ij}.$$

The natural parameter in (14) is $\theta_{ij} = (2\mu_{ij}^2)^{-1}$ and hence if θ_{ij} is written as the negative of the right hand side in (12), then obviously

$$\mu_{ij} = (2\theta_{ij})^{-1/2} = e^{-\frac{1}{2}\ln 2 + \frac{1}{2}\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}_{ij}'\boldsymbol{b}_i}, \qquad (15)$$

and once again the log link considerations can be invoked.

D. Details on Estimation and Use of the SAS Procedure NLMIXED

D.1 Overview of Estimation Strategies

The connections drawn out between GLMM, CM, MMM, COMMM, and bridge models have implications for their respective estimation strategies. In this section, we argue that all of the approaches considered can be fitted rather easily with standard software tools. Even in the binary case with conditional logit specification, all of the models can be fitted using a flexible parameter estimation device for non-linear mixed-effects models such as, for example, the SAS procedure NLMIXED. Parameter estimation and corresponding software tools have received considerable attention and Griswold and Zeger (2004) offer an extensive review. We mention, in particular, Pinheiro and Bates (2000), Diggle et al. (2002), Goldstein (2002), Skrondal and Rabe-Hesketh (2004), and Molenberghs and Verbeke (2005).

The most straightforward case is the standard GLMM, for which a variety of numerical methods to evaluate the marginal function have been proposed, such as first- and second-order Taylor series expansions (PQL, MQL, PQL2, MQL2), Laplace transformation based methods, and numerical integration (Gaussian quadrature and adaptive Gaussian quadrature). A review is given in Molenberghs and Verbeke (2005). Software implementations encompass the SAS procedures GLIMMIX and NLMIXED. While GLIMMIX is very stable and fast for GLMM, NLMIXED is convenient for all other cases. Also Monte Carlo based integration can be used. For the CM, Molenberghs, Verbeke, and Demétrio (2007) and Molenberghs et al. (2010) integrated analytically over the overdispersion random effects but not over the normal random effects, thus allowing for the use of NLMIXED. Heagerty (1999) and Griswold and Zeger (2004) also used NLMIXED, to implement the MMM, where the connector function follows from the appropriate integral. Evidently, this is most complex for the logit situation, owing to the lack of a closed-form solution. For the COMMM, Iddi and Molenberghs (2011), and Efendi, Molenberghs, and Iddi (2013) combined the CM and MMM manipulations to still allow for the use of NLMIXED. Finally, Wang and Louis (2003) transformed the logistic bridge random effect to a normal random effect. This is related to, but different from, the transformation method described in Liu and Yu (2007). Whereas here the transformation is in terms of cumulative densities, Liu and Yu (2007) propose a method in terms of densities, which could be useful in cases where a particular bridge density is easier to obtain than its corresponding cumulative distribution function. Wang and Louis (2004) extended their estimation method to allow for covariate-dependent random-intercept models.

D.2 Example Software Code

In this section, code is provided for the various analysis done on the toenail dataset. All are based on the SAS procedure NLMIXED with a few instances of the SAS procedure GLIMMIX as well.

```
/* data set to create auxiliary variables */
data test;
set m.toenail;
timeclss=time;
onyresp=y;
run;


proc glimmix data=test method=gauss(qpoints=20);
title 'regular GLMM with GLIMMIX, using adaptive Gaussian quadrature, logit';
class idnum timeclss;
model onyresp = treatn time treatn*time / dist=b solution;
random intercept / subject=idnum;
run;
```

```
proc nlmixed data=test qpoints=20;
title 'regular GLMM with NLMIXED, using adaptive Gaussian quadrature, logit';
parms beta0=-1.63 beta1=0.11 beta2=-0.4 beta3=0.16 d=4;
linpred = beta0+b+beta1*treatn + beta2*time + beta3*treatn*time;
p = exp(linpred)/(1+exp(linpred));
model onyresp ~ binary(p);
random b ~ normal(0,d*d) subject=idnum;
run;

proc nlmixed data=test qpoints=20;
title 'bridge model, using adaptive Gaussian quadrature, logit';
parms beta0=-1.63 beta1=0.11 beta2=-0.4 beta3=0.16 d=4;
pi = 3.1415926535897931;
uni = probnorm(b/d);
phi = 1.0/sqrt(1+3/pi/pi/d/d);
bridgel = 1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));
linpred = beta0+bridgel+beta1*treatn + beta2*time + beta3*treatn*time;
p = exp(linpred)/(1+exp(linpred));
model onyresp ~ binary(p);
random b ~ normal(0,d*d) subject=idnum;
run;

proc nlmixed data=test qpoints=10;
title 'standard combined model, using adaptive Gaussian quadrature, logit';
parms beta0=-1.54 beta1=-6.49 beta2=-16.27 beta3=-8.11 d=61 ratio=0.3
      /*beta=0.1 to 1 by 0.1 */;
teta = beta0+b+beta1*treatn + beta2*time + beta3*treatn*time;
expteta = exp(teta);
k = expteta/(1+expteta);
title'ratio=beta/alpha=meeschatten, startwaarde=0.3';
p = k/(1+ratio);
if onyresp=1 then ll=log(p);
if onyresp=0 then ll=log(1-p);
model onyresp ~ general(ll);
random b ~ normal(0,d*d) subject=idnum;
run;

proc glimmix data=test method=gauss(qpoints=20);
title 'regular GLMM with GLIMMIX, adaptive Gaussian quadrature, probit';
class idnum timeclss;
model onyresp = treatn time treatn*time
                / dist=b solution link=probit;
random intercept / subject=idnum;
run;

proc nlmixed data=test qpoints=50;
title 'regular GLMM with NLMIXED, adaptive Gaussian quadrature, probit';
parms beta0=-1.63 beta1=0.11 beta2=-0.4 beta3=0.16 d=4;
linpred = beta0 + b + beta1*treatn + beta2*time + beta3*treatn*time;
p = probnorm(linpred);
model onyresp ~ binary(p);
random b ~ normal(0,d*d) subject=idnum;
run;

proc nlmixed data=test qpoints=50;
```

```
title 'MMM with NLMIXED, adaptive Gaussian quadrature, probit';
parms beta0=-1.63 beta1=0.11 beta2=-0.4 beta3=0.16 d=2;
linpred = b+sqrt(1+d*d)*(beta0+beta1*treatn + beta2*time + beta3*treatn*time);
p = probnorm(linpred);
model onyresp ˜ binary(p);
random b ˜ normal(0,d*d) subject=idnum;
run;
```

Additional References

Efendi, A., Molenberghs, G., & Iddi, S. (2013). A marginalized combined model for repeated, overdispersed time-to-event outcomes. *Communication in Statistics-Theory and Methods*, accepted.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical Distributions*. New York: John Wiley & Sons.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Boca Raton: Chapman & Hall/CRC. http://dx.doi.org/10.1201/9781420011371

Liu, L., & Yu, Z. (2007). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine, 27*, 3105-3124. http://dx.doi.org/10.1002/sim.3153

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-Plus*. New-York: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4419-0318-1

Spiegel, M. R. (1968). *Mathematical Handbook of Formulas and Tables*. Schaum's Outline Series in Mathematics. New York: McGraw-Hill.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. London: Chapman & Hall/CRC. http://dx.doi.org/10.1201/9780203489437