Statistical Science
2010, Vol. 25, No. 3, 325–347
DOI: 10.1214/10-STS328

A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects

Geert Molenberghs, Geert Verbeke, Clarice G. B. Demétrio and Afrânio M. C. Vieira

Abstract. Non-Gaussian outcomes are often modeled using members of the so-called exponential family. Notorious members are the Bernoulli model for binary data, leading to logistic regression, and the Poisson model for count data, leading to Poisson regression. Two of the main reasons for extending this family are (1) the occurrence of overdispersion, meaning that the variability in the data is not adequately described by the models, which often exhibit a prescribed mean–variance link, and (2) the accommodation of hierarchical structure in the data, stemming from clustering in the data which, in turn, may result from repeatedly measuring the outcome, for various members of the same family, etc. The first issue is dealt with through a variety of overdispersion models, such as, for example, the beta-binomial model for grouped binary data and the negative-binomial model for counts. Clustering is often accommodated through the inclusion of random subject-specific effects. Though not always, one conventionally assumes such random effects to be normally distributed. While both of these phenomena may occur simultaneously, models combining them are uncommon. This paper proposes a broad class of generalized linear models accommodating overdispersion and clustering through two separate sets of random effects. We place particular emphasis on so-called 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 our family is more general. The binary, count and time-to-event cases are given particular emphasis. Apart from model formulation, we present an overview of estimation methods, and then settle for maximum likelihood estimation with analytic–numerical integration. Implications for the derivation of marginal correlations functions are discussed. The methodology is applied to data from a study in epileptic seizures, a clinical trial in toenail infection named onychomycosis and survival data in children with asthma.

Key words and phrases: Bernoulli model, Beta–binomial model, Cauchy distribution, conjugacy maximum likelihood, frailty model, negative-binomial model, Poisson model, strong conjugacy, Weibull model.

Geert Molenberghs is Professor of Biostatistics, I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium and I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium (e-mail: geert.molenberghs@uhasselt.be). Geert Verbeke is Professor of Biostatistics, I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium and I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium (e-mail: geert.verbeke@med.kuleuven.be). Clarice G. B. Demétrio is Professor of Biometry, ESALQ, Piracicaba, Saõ Paulo, Brazil (e-mail: clarice@esalq.usp.br). Afrânio M. C. Vieira is Professor of Statistics, ESALQ, Piracicaba, Saõ Paulo, Brazil and Universidade de Brasília, Department de Estatística, Brasília/DF, Brasil (e-mail: fran.usp@gmail.com).

1. INTRODUCTION

Next to continuous outcomes, binary and binomial outcomes, counts and times to event take a prominent place in applied modeling and the corresponding methodological literature. It is common to place such models within the generalized linear modeling (GLM) framework (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Agresti, 2002). This framework allows one to restrict specification to first and second moments only, on the one hand, or to fully formulate distributional assumptions, on the other hand. When the latter route is chosen, the exponential family (McCullagh and Nelder, 1989) provides an elegant and encompassing mathematical framework, because it has the normal, Bernoulli/binomial, Poisson and Weibull/exponential models as prominent members.

The elegance of the framework draws from certain linearity properties of the log-likelihood function, producing mathematically convenient score equations and ultimately convenient-in-use inferential instruments, both in terms of point and interval estimation as well as for hypothesis testing.

Nevertheless, it has been clear for several decades, for binomial, count and time-to-event data, that a key feature of the GLM framework and many of the exponential family members, the so-called *mean–variance relationship*, may be overly restrictive. By this relationship, we indicate that the variance is a deterministic function of the mean. For example, for Bernoulli outcomes with success probability $\mu = \pi$, the variance is $v(\mu) = \pi(1 - \pi)$, for counts $v(\mu) = \mu$ and for the exponential model $v(\mu) = \mu^2$. In contrast, for continuous, normally distributed outcomes, the mean and variance are entirely separate parameters. While i.i.d. binary data cannot contradict the mean–variance relationship, i.i.d. binomial data, counts and survival data can. This explains why early work has been devoted to formulating models that explicitly allow for overdispersion or, more generally, to proposing models that enjoy less restrictive mean–variance relationships. For purely binary data, hierarchies need to be present in the data in order to violate the mean–variance link. One such class of hierarchies is with repeated measures or longitudinal data, where an outcome on a study subject is recorded repeatedly over time. With such models gaining momentum, not only for the Gaussian case (Verbeke and Molenberghs, 2000), but also for non-Gaussian data (Molenberghs and Verbeke, 2005), extensions of the GLM framework have been formulated. For other types of outcomes, such hierarchical settings further compound the issue of overly restrictive variance relationships. In all cases, hierarchies induce association. These features taken together call for very flexible models, doing proper justice to each of the mean, variance and association structures.

Hinde and Demétrio (1998a, 1998b) provide broad overviews of approaches for dealing with overdispersion, considering moment-based as well as full-distribution avenues. Placing most emphasis on the binomial and Poisson settings, they pay particular attention to random-effects-based solutions to the problem, including but not limited to the beta-binomial model (Skellam, 1948; Kleinman, 1973) for binary and binomial data and with beta random effects, and the negative-binomial model (Breslow, 1984; Lawless, 1987), where the natural parameter is assumed to follow a gamma distribution. The said gamma distribution also features in many so-called frailty models, that is, specific random-effects models for time-to-event data (Duchateau and Janssen, 2007). On the other hand, especially focusing on hierarchical data, the so-called generalized linear mixed model (GLMM, Engel and Keen, 1994; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993) has gained popularity as a tool to accommodate overdispersion and/or hierarchy-induced association for outcomes that are not necessarily of a Gaussian type, in spite of problems, not only of a computational type, but also in terms of interpretation. These arise from the combination of general exponential family models with normally distributed random effects. Unlike for Gaussian data, the derivation of marginal moments and joint distributions is less than straightforward, even though in this paper we make progress beyond what is available in the literature. Part of GLMMs popularity originates from the availability of implementations in a variety of standard software packages. Other solutions to accommodating overdispersion include mixture modeling and specific models for zero-inflated Poisson models (Ridout, Demétrio and Hinde, 1998; Böhning, 2000; McLachlan and Peel, 2000).

Important unifying and computational progress has been made by Lee and Nelder (1996, 2001a, 2001b, 2003) (see also Lee, Nelder and Pawitan, 2006) by proposing so-called *hierarchical generalized linear models*, offering a broad class of outcome and random-effects distributions, combined with appealing computational schemes. Unification has also been reached by Skrondal and Rabe-Hesketh (2004), who assemble under the same roof a number of modeling strands, such

as multilevel modeling, structural equations modeling, latent variables, latent classes and random-effects models for longitudinal and otherwise hierarchical data.

In this paper we introduce a general and flexible framework for such combinations, starting from arbitrary generalized linear models and exponential family members. Specific emphasis is placed on normally distributed, binary, binomial, count and time-to-event outcomes. There are various reasons to do so. First, non-Gaussian hierarchical data exhibit three important features: (1) the mean structure; (2) the variance structure; and (3) the correlation structure. Our proposed framework features: (a) a mean structure; (b) overdispersion, often conjugate random-effects; (c) normal random effects. It will be clear from our case studies that model fit can be improved, and hence model interpretation changed, by shifting to the extended model. Second, especially in cases where the variance and/or correlation structures are of interest (e.g., surrogate marker evaluation, psychometric evaluation, etc.), such extensions are useful. Third, even when interest remains with more conventional models, such as the GLMM, the extended model can serve as a goodness-of-fit tool. Fourth, because we can derive closed-form expressions for both standard and extended models, the accuracy of parameter estimation and resulting inferences can be improved, while obviating the need for tedious numerical integration techniques. Fifth, the analysis of the case studies corroborates this need. Such needs were recognized by Booth et al. (2003) and Molenberghs, Verbeke and Demétrio (2007) who, in the context of count data, formulated a model combining normal and gamma random effects.

The paper is organized as follows. In Section 2 three motivating case studies, with binary data, counts and survival data, respectively, are described, with analyses reported near the end of the manuscript, in Section 6. Basic ingredients for our modeling framework, standard generalized linear models, extensions for overdispersion and the generalized linear mixed model, are the subject of Section 3. The proposed, combined model is described and further studied in Section 4. Avenues for parameter estimation and ensuing inferences are explored in Section 5. There are several appendices. Supplementary Material A offers generic approximations for means and variances. Supplementary Material B–E provide details for the Poisson case, the binary case with logit link and the binary case with probit link, and the time-to-event case, respectively. Implications of our findings for the derivation of marginal correlation functions are the topic of Section F in the Supplementary Material.

2. CASE STUDIES

We will describe three case studies. The first one producing count data, the second one with binary data, and the third one of a time-to-event type.

2.1 A Clinical Trial in Epileptic Patients

The data considered here are obtained 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. The study is described in full detail in Faught et al. (1996). The randomization of epilepsy patients took place after a 12-week baseline period that served as a stabilization period for the use of AED's, and during which the number of seizures were counted. After that period, 45 patients were assigned to the placebo group and 44 to the active (new) treatment group. Patients were then measured weekly. Patients were followed (double-blind) during 16 weeks, after which they were entered into 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 most recent week. The research question is whether or not the additional new treatment reduces the number of epileptic seizures.

2.2 A Casfe Study in Onychomycosis

These data come from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (coded as A and B) for toenail dermatophyte onychomycosis (TDO), described in full detail by De Backer et al. (1996). TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons (Roberts et al., 1992). Anti-fungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthy. The development of new such compounds, however, 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, distributed over 36 centers, were randomized. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment and every 3 months afterward, resulting in a maximum of 7 measurements per subject. At the first occasion, the treating physician indicates one of the affected toenails

as the target nail, the nail which will be followed over time. We will restrict our analyses to only those patients for which the target nail was one of the two big toenails (146 and 148 subjects, in group A and group B, respectively). One of the responses of interest was the unaffected nail length, measured from the nail bed to the infected part of the nail, which is always at the free end of the nail, expressed in mm. This outcome has been studied extensively in Verbeke and Molenberghs (2000). Another important outcome in this study 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.

2.3 Recurrent Asthma Attacks in Children

These data have been studied in Duchateau and Janssen (2007). Asthma is occurring more and more frequently in very young children (between 6 and 24 months). Therefore, a new application of an existing anti-allergic drug is administered to children who are at higher risk to develop asthma in order to prevent it. A prevention trial is set up with such children randomized to placebo or drug, and the asthma events that developed over time are recorded in a diary. Typically, a patient has more than one asthma event. The different events are thus clustered within a patient and ordered in time. This ordering can be taken into account in the model. The data are presented in calendar time format, where the time at risk for a particular event is the time from the end of the previous event (asthma attack) to the start of the next event (start of the next asthma attack). A particular patient has different periods at risk during the total observation period which are separated either by an asthmatic event that lasts one or more days or by a period in which the patient was not under observation. The start and end of each such risk period is required, together with the status indicator to denote whether the end of the risk period corresponds to an asthma attack or not.

3. REVIEW OF KEY INGREDIENTS

In Section 3.1 we will first describe the conventional exponential family and generalized linear modeling based on it. Section 3.2 is devoted to a brief review of models for overdispersion. Section 3.3 focuses on the normally distributed case.

3.1 Standard Generalized Linear Models

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

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

for a specific set of unknown parameters η and ϕ, and for known functions $\psi(\cdot)$ and $c(\cdot,\cdot)$. Often, η and ϕ are termed "natural parameter" (or "canonical parameter") and "dispersion parameter," respectively.

It can easily be shown (Molenberghs and Verbeke, 2005) that the first two moments follow from the function $\psi(\cdot)$ as

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

$$\mathrm{Var}(Y) = \sigma^2 = \phi\psi''(\eta). \tag{3}$$

An important implication is that, in general, the mean and variance are related through $\sigma^2 = \phi\psi''[\psi'^{-1}(\mu)] = \phi v(\mu)$, with $v(\cdot)$ the so-called variance function, describing the mean–variance relationship.

Key instances of the exponential family for normal, binary, count and time-to-event data are listed in Table 1, along with their exponential family elements. The normal model is special, in particular, also because the overdispersion parameter is needed to allow for a variance other than unity. As a result, the mean–variance relationship is absent for this model, but present for all others. In the binary case, an alternative to the Bernoulli model with logit link is the probit model, where $\eta = \Phi^{-1}(\pi)$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Evidently, this model is slightly less standard because the probit model is not the natural link, as we will see in Section 4.6, it has appeal in the overdispersed and/or repeated contexts.

In the Weibull and exponential model, the decomposition $\varphi = \lambda e^{\mu}$ is often employed, with notation as in Table 1, allowing for μ to become a function of covariates. Evidently, here, while μ is a component of the mean function, it is in itself not equal to the mean. Note also that the Weibull model does not belong to the exponential family in a conventional sense, unless in a somewhat contrived fashion where y is replaced by y^{ρ}. In the mean and variance expressions for the Weibull (Table 1), $\Gamma(\cdot)$ represents the gamma function.

In some situations, for example, when quasi-likelihood methods are employed (McCullagh and Nelder, 1989; Molenberghs and Verbeke, 2005), no full distributional assumptions are made, but one rather restricts

TABLE 1

Conventional exponential family members and extensions with conjugate random effects

Element	**Notation**	**Continuous**	**Binary**	**Count**	**Time to event**	
Standard univariate exponential family						
Model		Normal	Bernoulli	Poisson	Exponential	Weibull
Model	$f(y)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\mu)^2/(2\sigma^2)}$	$\pi^y(1-\pi)^{1-y}$	$\frac{e^{-\lambda}\lambda^y}{y!}$	$\varphi e^{-\varphi y}$	$\varphi\rho y^{\rho-1}e^{-\varphi y^\rho}$
Nat. param	η	μ	$\ln[\pi/(1-\pi)]$	$\ln\lambda$	$-\varphi$	
Mean function	$\psi(\eta)$	$\eta^2/2$	$\ln[1+\exp(\eta)]$	$\lambda=\exp(\eta)$	$-\ln(-\eta)$	
Norm. constant	$c(y,\phi)$	$\frac{\ln(2\pi\phi)}{2}-\frac{y^2}{2\phi}$	0	$-\ln y!$	0	
(Over)dispersion	ϕ	σ^2	1	1	1	
Mean	μ	μ	π	λ	φ^{-1}	$\varphi^{-1/\rho}\Gamma(\rho^{-1}+1)$
Variance	$\phi v(\mu)$	σ^2	$\pi(1-\pi)$	λ	φ^{-2}	$\varphi^{-2/\rho}[\Gamma(2\rho^{-1}+1)-\Gamma(\rho^{-1}+1)^2]$
Exponential family with conjugate random effects						
Model		Normal–normal	Beta–binomial	Negative-binomial	Exponential–gamma	Weibull–gamma
Hier. model	$f(y\|\theta)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\theta)^2/(2\sigma^2)}$	$\theta^y(1-\theta)^{1-y}$	$\frac{e^{-\theta}\theta^y}{y!}$	$\varphi\theta e^{-\varphi\theta y}$	$\varphi\theta\rho y^{\rho-1}e^{-\varphi\theta y^\rho}$
RE model	$f(\theta)$	$\frac{1}{\sqrt{d}\sqrt{2\pi}}e^{-(\theta-\mu)^2/(2d)}$	$\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$
Marg. model	$f(y)$	$\frac{1}{\sqrt{\sigma^2+d}\sqrt{2\pi}}e^{-(y-\mu)^2/(2(\sigma^2+d))}$	$(\alpha+\beta)\frac{\Gamma(\alpha)}{\Gamma(\alpha+y)}\frac{\Gamma(\beta)}{\Gamma(\beta+1-y)}$	$\frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)}(\frac{\beta}{\beta+1})^y(\frac{1}{\beta+1})^\alpha$	$\frac{\varphi\alpha\beta}{(1+\varphi\beta y)^{\alpha+1}}$	$\frac{\varphi\rho y^{\rho-1}\alpha\beta}{(1+\varphi\beta y^\rho)^{\alpha+1}}$
	$h(\theta)$	θ	$\ln[\theta/(1-\theta)]$	$\ln(\theta)$	$-\theta$	$-\theta$
	$g(\theta)$	$-\frac{1}{2}\theta^2$	$-\ln(1-\theta)$	θ	$-\ln(\theta)/\varphi$	$-\ln(\theta)/\varphi$
	ϕ	σ^2	1	1	$1/\varphi$	$1/\varphi$
	γ	$1/d$	$\alpha+\beta-2$	$1/\beta$	$\varphi(\alpha-1)$	$\varphi(\alpha-1)$
	ψ	μ	$\frac{\alpha-1}{\alpha+\beta-2}$	$\beta(\alpha-1)$	$[\beta\varphi(\alpha-1)]^{-1}$	$[\beta\varphi(\alpha-1)]^{-1}$
	$c(y,\phi)$	$-\frac{1}{2}\phi y^2-\frac{1}{2}\ln(\frac{2\pi}{\phi})$	0	$-\ln(y!)$	$\ln(\varphi)$	$\ln(\varphi\rho y^{\rho-1})$
	$c^*(\gamma,\psi)$	$-\frac{1}{2}\gamma\psi^2-\frac{1}{2}\ln(\frac{2\pi}{\gamma})$	$-\ln B(\gamma\psi+1,\ \gamma-\psi\gamma+1)$	$(1+\gamma\psi)\ln\gamma-\ln\Gamma(1+\gamma\psi)$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi)-\ln\Gamma(\frac{\gamma+\varphi}{\varphi})$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi)-\ln\Gamma(\frac{\gamma+\varphi}{\varphi})$
Mean	$E(Y)$	μ	$\frac{\alpha}{\alpha+\beta}$	$\alpha\beta$	$[\varphi(\alpha-1)\beta]^{-1}$	$\frac{\Gamma(\alpha-\rho^{-1})\Gamma(\rho^{-1}+1)}{(\varphi\beta)^{1/\rho}\Gamma(\alpha)}$
Variance	$\mathrm{Var}(Y)$	σ^2+d	$\frac{\alpha\beta}{(\alpha+\beta)^2}$	$\alpha\beta(\beta+1)$	$\alpha[\varphi^2(\alpha-1)^2(\alpha-2)\beta^2]^{-1}$	$\frac{1}{\rho(\varphi\beta)^{2/\rho}\Gamma(\alpha)}[2\Gamma(\alpha-2\rho^{-1})\Gamma(2\rho^{-1})-\frac{\Gamma(\alpha-\rho^{-1})^2\Gamma(\rho^{-1})^2}{\rho\Gamma(\alpha)}]$

to specifying the first two moments (2) and (3). In such an instance, the variance function $v(\mu)$ can be chosen in accordance with a particular member of the exponential family. If not, then parameters cannot be estimated using maximum likelihood principles. Instead, a set of estimating equations needs to be specified, the solution of which is referred to as the quasi-likelihood estimates.

In a regression context, where one wishes to explain variability between outcome values based on measured covariate values, the model needs to incorporate covariates. This leads to so-called generalized linear models. Let $Y_1, \ldots, Y_N$ be a set of independent outcomes, and let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ represent the corresponding p-dimensional vectors of covariate values. It is assumed that all Y_i have densities $f(y_i|\eta_i, \phi)$, which belong to the exponential family, but a different natural parameter η_i is allowed per observation. Specification of the generalized linear model is completed by modeling the means μ_i as functions of the covariate values. More specifically, it is assumed that $\mu_i = h(\eta_i) = h(\mathbf{x}_i'\boldsymbol{\xi})$, for a known function $h(\cdot)$, and with $\boldsymbol{\xi}$ a vector of p fixed, unknown regression coefficients. Usually, $h^{-1}(\cdot)$ is called the link function. In most applications, the so-called natural link function is used, that is, $h(\cdot) = \psi'(\cdot)$, which is equivalent to assuming $\eta_i = \mathbf{x}_i'\boldsymbol{\xi}$. Hence, it is assumed that the natural parameter satisfies a linear regression model.

3.2 Overdispersion Models

It is clear from Table 1 that the standard Bernoulli, Poisson and exponential models force the mean and variance functions to depend on a single parameter. However, comparing the sample average with the sample variance might already reveal in certain applications that this assumption is not in line with a particular set of data, for count and time-to-event data, for example. While this is one of the senses in which the binary case is somewhat exceptional, because a set of i.i.d. Bernoulli data cannot contradict the mean–variance relationship, it would still hold for the related binomial case, where the data take the form of n_i successes out of z_i trials.

Therefore, a number of extensions have been proposed, as briefly mentioned in the Introduction. Hinde and Demétrio (1998a, 1998b) provide general treatments of overdispersion. The Poisson case received particular attention by Breslow (1984) and Lawless (1987). Molenberghs and Verbeke (2005) mention various model-based approaches that accommodate overdispersion, including the beta-binomial model (Skellam, 1948), the Bahadur model (1961), the multivariate probit model (Dale, 1986; Molenberghs and Lesaffre, 1994) and certain versions of the generalized linear mixed model (Breslow and Clayton, 1993). The latter family will be studied in Section 3.3.

A straightforward and commonly encountered step is to allow the overdispersion parameter $\phi \neq 1$, so that (3) produces $\mathrm{Var}(Y) = \phi v(\mu)$. This is in line with the moment-based approach mentioned in the previous section, but can also be engendered by fully parametric assumptions.

An elegant way forward is through a two-stage approach. For binary data, one would assume that $Y_i|\pi_i \sim \mathrm{Bernoulli}(\pi_i)$ and further that π_i is a random variable with $\mathrm{E}(\pi_i) = \mu_i$ and $\mathrm{Var}(\pi_i) = \sigma_i^2$. Using iterated expectations, it follows that

$$\begin{aligned} E(Y_i) &= E[E(Y_i|\pi_i)] = E(\pi_i) = \mu_i, \\ \mathrm{Var}(Y_i) &= E[\mathrm{Var}(Y_i|\pi_i)] + \mathrm{Var}[E(Y_i|\pi_i)] \\ &= E[\pi_i(1-\pi_i)] + \mathrm{Var}(\pi_i) \\ &= E(\pi_i) - E(\pi_i^2) + E(\pi_i^2) - E(\pi_i)^2 \\ &= \mu_i(1-\mu_i), \end{aligned}$$

underscoring that purely Bernoulli data are unable to capture overdispersion.

Likewise, for the Poisson case, we assume that $Y_i|\zeta_i \sim \mathrm{Poi}(\zeta_i)$ and then that ζ_i is a random variable with $\mathrm{E}(\zeta_i) = \mu_i$ and $\mathrm{Var}(\zeta_i) = \sigma_i^2$. Also here then, it follows that

$$\begin{aligned} E(Y_i) &= E[E(Y_i|\zeta_i)] = E(\zeta_i) = \mu_i, \\ \mathrm{Var}(Y_i) &= E[\mathrm{Var}(Y_i|\zeta_i)] + \mathrm{Var}[E(Y_i|\zeta_i)] \\ &= E(\zeta_i) + \mathrm{Var}(\zeta_i) = \mu_i + \sigma_i^2. \end{aligned}$$

Note that we have not assumed a particular distributional form for the random effects π_i and ζ_i, respectively. Hence, this gives rise to a semi-parametric specification. Similar routes can be followed for other GLM, too.

In case it is considered advantageous to make full distributional assumptions about the random effects, common choices are the beta distribution for π_i and the gamma distribution for ζ_i; of course, these are not the only ones.

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:

$$f(y_i) = \int f(y_i|\theta_i) f(\theta_i)\, d\theta_i. \tag{4}$$

It is easy to extend this model to the case of repeated measurements. We then assume a hierarchical data structure, where now Y_{ij} denotes the jth outcome measured for cluster (subject) i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and $\mathbf{Y}_i$ is the n_i-dimensional vector of all measurements available for cluster i. In the repeated-measures case, the scalar ζ_i becomes a vector $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{in_i})'$, with $E(\boldsymbol{\zeta}_i) = \boldsymbol{\mu}_i$ and $\mathrm{Var}(\boldsymbol{\zeta}_i) = \boldsymbol{\Sigma}_i$. For example, for the Poisson case, similar logic as in the univariate case produces $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$ and $\mathrm{Var}(\mathbf{Y}_i) = M_i + \Sigma_i$, where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the diagonal. Note that a diagonal structure of M_i reflects the conditional independence assumption: all dependence between measurements on the same unit stems from the random effects. Generally, a versatile class of models results. For example, assuming that the components of $\boldsymbol{\zeta}_i$ are independent, a pure overdispersion model follows, without correlation between the repeated measures. On the other hand, assuming $\zeta_{ij} = \zeta_i$, that is, that all components are equal, then $\mathrm{Var}(\mathbf{Y}_i) = M_i + \sigma_i^2 J_{n_i}$, where J_{n_i} is an $(n_i \times n_i)$-dimensional matrix of ones. Such a structure can be seen as a general version of compound symmetry. Of course, one can also combine general correlation structures between the components of $\boldsymbol{\zeta}_i$.

Alternatively, this repeated version of the overdispersion model can be combined with normal random effects in the linear predictor. This very specific choice, proposed also by Thall and Vail (1990) and Dean (1991), for the count case, will be the focus of the next section.

General marginalization (4) may seem an elegant and general principle, there is the issue of having to decide which parameter to turn into a random one. This is especially true if one considers the need to select an actual distributional form for the random effect. A noteworthy exception is, as always, the linear mixed model, combining a normal hierarchical model with a normal random effect. It forms the basis of the two strands of random-effects models that are potentially brought together in the combined models of Section 4: on the one hand, normal random effects can be considered with nonnormal outcomes, producing the GLMM; on the other hand, gamma random effects for the Poisson model, beta random effects with binomial data and gamma random effects for the Weibull model can be considered. This is, seemingly, a disparate collection. However, they are bound together by the property of *conjugacy*, in the sense of Cox and Hinkley (1974), page 370, and Lee, Nelder and Pawitan (2006), page 178. The topic is also discussed by Agresti (2002). Informally, conjugacy refers to the fact that the hierarchical and random-effects densities have similar algebraic forms. Conjugate distributions produce a general and closed-form solution for the corresponding marginal distribution.

We will first define conjugacy as is conventionally done, that is, in models without the normal random effects and then, in Section 4, introduce a further property, *strong conjugacy*, necessary for situations where both normal and conventional conjugate random effects are present. To simplify notation, we will provide the definition at a general distribution level, with neither subject- nor measurement-specific subscripts, so that it can be applied to both univariate and longitudinal data. The hierarchical and random-effects densities are said to be conjugate if and only if they can be written in the generic forms

$$f(y|\theta) = \exp\{\phi^{-1}[yh(\theta) - g(\theta)] + c(y, \phi)\}, \tag{5}$$

$$f(\theta) = \exp\{\gamma[\psi h(\theta) - g(\theta)] + c^*(\gamma, \psi)\}, \tag{6}$$

where $g(\theta)$ and $h(\theta)$ are functions, ϕ, γ and ψ are parameters, and the additional functions $c(y, \phi)$ and $c^*(\gamma, \psi)$ are so-called normalizing constants. It can then be shown, upon constructing the joint distribution and then integrating over the random effect, that the marginal model resulting from (5) and (6) equals

$$f(y) = \exp\left[c(y, \phi) + c^*(\gamma, \psi) - c^*\left(\phi^{-1} + \gamma, \frac{\phi^{-1}y + \gamma\psi}{\phi^{-1} + \gamma}\right)\right]. \tag{7}$$

Table 1 gives model elements, such as density or probability mass functions, conditional on random effects and marginalized over these, as well as the random effects distributions. For all models considered, the constants and functions featuring in (5)–(6) are listed, and finally marginal means and variances are provided. For some models, these are well known (Hinde and Demétrio, 1998a, 1998b) and/or easy to derive. For the time-to-event models, a sketch can be found in Appendix E. While there, the focus is on the combined version of Section 4.8, the overdispersion case considered here follows as a special case.

In the case of binary data, the model in Table 1 is the familiar beta-binomial model. Note that the variance still obeys the usual Bernoulli variance structure. This is entirely natural, given that we still focus on a single binary outcome, in contrast to the more conventional binomial basis model, where data of the format "z_i successes out n_i trials" are considered. We do not

consider this situation in this section, but rather leave it to Section 4. In such a case, the variance structure becomes $\pi_i(1-\pi_i)[1+\rho_i(n_i-1)]$, where ρ_i is a measure for correlation. All parameters, p_i and ρ_i, can be expressed in terms of α_i and β_i, "cluster-specific" versions of the beta parameters.

For count data, the familiar negative-binomial model results. Unlike in the binary case, univariate counts are able to violate the mean–variance relationship of the Poisson distribution, hence the great popularity of this and other types of models for overdispersion. The same applies to the exponential distribution. Of course, already the Weibull model, with its extra parameter ρ, alleviates the constraint.

The normal distribution case is a special one. Not only is it self-conjugate, also the model is not identified, unlike all others. This is because both random terms, seen from writing $Y_i = \mu_i + b_i + \varepsilon_i$, are in direct, linear relationship with each other. In the generalized linear context, the various random terms have no direct linear alliance. The normal case will continue to be "the odd one out" in models to come (Sections 3.3 and 4).

The parameters α and β in the beta and gamma distributions are not always jointly identified. It is therefore customary to impose restrictions, such as setting one of them equal to a fixed value, for example, $\alpha = 1$, or constraining their mean or variance, etc. Such constraints operate differently, depending on other elements present in the models. For example, the presence of additional random effects in a model for repeated measures, such as in Section 4, alters the meaning and restrictiveness of such constraints.

Recall that the models at the bottom part of Table 1 are not the only options, but rather common, elegant choices, where the elegance draws to a large extent from conjugacy.

3.3 Models with Normal Random Effects

The generalized linear mixed model (Engel and Keen, 1994; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993) is likely the most frequently used random-effects model in the context of perhaps non-Gaussian repeated measurements. Not only is it a relatively straightforward extension of the generalized linear model for independent data (Section 3.1) to the context of hierarchically organized data, on the one hand, and the linear mixed model (Verbeke and Molenberghs, 2000), on the other hand, but there is also a wide range of software tools available for fitting such models.

Let Y_{ij} be the jth outcome measured for cluster (subject) $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and group the n_i measurements into a vector $\mathbf{Y}_i$. Assume that, in analogy with Section 3.1, conditionally upon q-dimensional random effects $\mathbf{b}_i \sim N(\mathbf{0}, D)$, the outcomes Y_{ij} are independent with densities of the form

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

with

$$\eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) = \eta[E(Y_{ij}|\mathbf{b}_i, \boldsymbol{\xi})] = \mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i \tag{9}$$

for a known link function $\eta(\cdot)$, with $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ p-dimensional and q-dimensional vectors of known covariate values, with $\boldsymbol{\xi}$ a p-dimensional vector of unknown fixed regression coefficients, and with ϕ a scale (overdispersion) parameter. Finally, let $f(\mathbf{b}_i|D)$ be the density of the $N(\mathbf{0}, D)$ distribution for the random effects $\mathbf{b}_i$.

These models closely follow the ones formulated in the top part of Table 1, with key differences that now: (1) data hierarchies are allowed for, in our setting owing to the longitudinal collection of data; (2) the natural parameter is written as a linear predictor, a function of both fixed and random effects.

Obviously, such models can be formulated for all data settings considered in Table 1 and beyond. This is conventionally done for continuous, Gaussian data, producing the linear mixed-effects model (Verbeke and Molenberghs, 2000), as well as for binary data and counts. This kind of model is a bit less common for survival data, where so-called frailty models (Duchateau and Janssen, 2007), rather of the type described in Section 3.2, are more standard. Of course, also the accelerated failure time model with random effects deserves mention, given that it takes the form of a linear mixed model for logarithmic time.

We will not consider explicit expressions for such models here, because they are relatively well studied (Fahrmeir and Tutz, 2001; Molenberghs and Verbeke, 2005) and, at any rate, conveniently follow as special cases from the combined models of Section 4.

4. MODELS COMBINING CONJUGATE AND NORMAL RANDOM EFFECTS

4.1 General Model Formulation

Integrating both the overdispersion effects of Table 1 (Section 3.2) as well as the normal random effects of

Section 3.3 into the generalized linear model framework produces the following general family:

$$f_i(y_{ij}|\mathbf{b}_i, \boldsymbol{\xi}, \theta_{ij}, \phi) = \exp\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\}, \tag{10}$$

with notation similar to the one used in (8), but now with conditional mean

$$E(Y_{ij}|\mathbf{b}_i, \boldsymbol{\xi}, \theta_{ij}) = \mu_{ij}^c = \theta_{ij}\kappa_{ij}, \tag{11}$$

where the random variable $\theta_{ij} \sim \mathcal{G}_{ij}(\vartheta_{ij}, \sigma_{ij}^2)$, $\kappa_{ij} = g(\mathbf{x}_{ij}'\boldsymbol{\xi} + \mathbf{z}_{ij}'\mathbf{b}_i)$, ϑ_{ij} is the mean of θ_{ij} and σ_{ij}^2 is the corresponding variance. Finally, as before, $\mathbf{b}_i \sim N(\mathbf{0}, D)$. Write $\eta_{ij} = \mathbf{x}_{ij}'\boldsymbol{\xi} + \mathbf{z}_{ij}'\mathbf{b}_i$. Unlike in Section 3.3, we now have two different notations, η_{ij} and λ_{ij}, to refer to the linear predictor and/or the natural parameter. The reason is that λ_{ij} encompasses the random variables θ_{ij}, whereas η_{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 $\mathbf{b}_i$, are independent of each other. Regarding the components θ_{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 $\mathcal{G}_{ij}(\vartheta_{ij}, \sigma_{ij}^2)$ needs to be replaced with a multivariate one; and (3) they are equal to each other, useful in applications with exchangeable outcomes Y_{ij}.

Obviously, parameterization (11) allows for random effects θ_{ij} capturing overdispersion, and formulated directly at mean scale, such as described in Section 3.2, whereas κ_{ij} could be considered the GLMM component, as in Section 3.3. The relationship between mean and natural parameter now is

$$\lambda_{ij} = h(\mu_{ij}^c) = h(\theta_{ij}\kappa_{ij}). \tag{12}$$

We can still apply standard GLM ideas, in particular, (2) and (3), to derive the mean and variance, combined with iterated-expectation-based calculations. For the mean, it follows that

$$E(Y_{ij}) = E(\theta_{ij})E(\kappa_{ij}) = E[h^{-1}(\lambda_{ij})]. \tag{13}$$

4.2 Generic Approximations for Marginal Model Elements

As we will see in ensuing specific cases (Sections 4.4–4.8), (13) allows for explicit expressions in a good number of cases. Generic mean, variance and covariance approximations can be derived using the expansion, around $\mathbf{b}_i = \mathbf{0}$,

$$\kappa_{ij} \approx g(\eta_{ij}) + g'(\eta_{ij})\mathbf{z}_{ij}'\mathbf{b}_i + \tfrac{1}{2}g''(\eta_{ij})\mathbf{z}_{ij}'\mathbf{b}_i\mathbf{b}_i'\mathbf{z}_{ij}.$$

Details and expressions are provided in Appendix A.

4.3 Strong Conjugacy

In Section 3.2 the concept of conjugacy was introduced and exemplified in a number of cases (see Table 1). It is of interest to explore under what conditions Model (10) still allows for conjugacy. The complication is the presence of the multiplicative factor κ_{ij} in the mean structure. To make progress, we will study how conjugacy plays out between Model (10) and the distribution of the random effect θ_{ij}, *given* the multiplicative factor κ_{ij}. In other words, conjugacy will be considered conditional upon the normally-distributed random effect $\mathbf{b}_i$. To this effect, write (suppressing nonessential arguments from the functions)

$$f(y|\kappa\theta) = \exp\{\phi^{-1}[yh(\kappa\theta) - g(\kappa\theta)] + c(y, \phi)\}, \tag{14}$$

generalizing (5), and retain (6). Applying the transformation theorem to (6) leads to

$$f(\theta|\gamma, \psi) = \kappa \cdot f(\kappa\theta|\widetilde{\gamma}, \widetilde{\psi}).$$

Next, we request that the parametric form (6) be maintained:

$$f(\kappa\theta) = \exp\{\gamma^*[\psi^* h(\kappa\theta) - g(\kappa\theta)] + c^{**}(\gamma^*, \psi^*)\}, \tag{15}$$

where the parameters γ^* and ψ^* follow from $\widetilde{\gamma}$ and $\widetilde{\psi}$ upon absorption of κ. Then, the marginal model, in analogy with (7), equals

$$f(y|\kappa) = \exp\left\{c(y, \phi) + c^{**}(\gamma^*, \psi^*) + c^{**}\left(\phi^{-1} + \gamma^*, \frac{\phi^{-1}y + \gamma^*\psi^*}{\phi^{-1} + \gamma^*}\right)\right\}. \tag{16}$$

Evidently, not every model satisfying conjugacy in the sense of Section 3.2 will allow for the present form of conjugacy. We will refer to this condition as *strong conjugacy*. Examples include the normal, Poisson and Weibull (and hence exponential) models with normal, gamma and gamma random effects, respectively. A counterexample is provided by the Bernoulli, and hence also binomial, model. Because the probit model does not allow for conjugacy, not even in the usual sense, it is out of the picture here, too. The latter does not preclude the existence of closed forms in the probit case, as we will see in Section 4.7.

Note that the transition from strong conjugacy is a property entirely of the random-effects distribution, and not of the data model, the latter of which is needed, of course, for conjugacy itself. For example,

for gamma random effects, we can write

$$\begin{aligned}\frac{1}{\kappa} f(\theta|\alpha,\beta) &= \frac{1}{\kappa}\frac{1}{\beta^\alpha\Gamma(\alpha)}\theta^{\alpha-1}e^{-\theta/\beta}\\ &= \frac{1}{(\kappa\beta)^\alpha\Gamma(\alpha)}(\kappa\theta)^{\alpha-1}e^{-(\kappa\theta)/(\kappa\beta)} \qquad (17)\\ &= f(\kappa\beta|\alpha,\kappa\beta)\end{aligned}$$

and, hence, a scaled version of a gamma random effect is still a gamma random effect, with retention of α and rescaling of β to $\kappa\beta$.

The importance of strong conjugacy lies, among others, in the easy integration over the nonnormal random effects θ_{ij}. As a consequence, the resulting density is conditional on κ and hence on $\mathbf{b}_i$ only, implying that standard software for generalized linear or nonlinear mixed-effects models, such as the SAS procedure NLMIXED, can be employed, a point to which we will return in Section 5.

We will now consider the normal, Poisson, binary and time-to-event cases in turn. Details of the calculations for the Poisson case are given in Molenberghs, Verbeke and Demétrio (2007) and summarized in Appendix B, while the binary and time-to-event cases are supported by Appendices C, D and E, respectively.

There is no need to spell out the various models in detail. The different versions of (10) follow straightforwardly upon combining the models formulated in Table 1 with the GLMM (8) and corresponding linear predictor (9). Precisely, the effect θ ought to be replaced by $\theta_{ij}\kappa_{ij}$, where κ_{ij} is defined by setting $\eta = \eta_{ij}$ equal to the linear predictor whence κ_{ij} is expressed, for the respective models, as μ, π, λ and ϕ.

4.4 Specific Case: Continuous, Normally Distributed Data

The fully hierarchically specified linear mixed-effects model takes the form (Verbeke and Molenberghs, 2000)

$$\mathbf{Y}_i|\mathbf{b}_i \sim N(X_i\boldsymbol{\xi} + Z_i\mathbf{b}_i, \Sigma_i), \qquad (18)$$

$$\mathbf{b}_i \sim N(0, D), \qquad (19)$$

where $\boldsymbol{\xi}$ is a vector of fixed effects, and X_i and Z_i are design matrices. The rows of $X_i\boldsymbol{\xi} + Z_i\mathbf{b}_i$ are made up by the linear predictors (9).

Based upon (18) and (19), the marginal model can be derived:

$$\mathbf{Y}_i \sim N(X_i\boldsymbol{\xi}, V_i = Z_i D Z_i' + \Sigma_i). \qquad (20)$$

We evidently consider a single set of random effects only, because, in this case, the normal and conjugate random effects coincide, a unique feature of the normal model. Strong conjugacy is a fortiori evident.

4.5 Specific Case: Poisson-Type Models for Count Data

From the general developments above, the Poisson model with gamma and normal random effects combined naturally follows. By way of overview, let us assemble all model elements:

$$Y_{ij} \sim \text{Poi}(\theta_{ij}\kappa_{ij}), \qquad (21)$$

$$\kappa_{ij} = \exp(\mathbf{x}_{ij}'\boldsymbol{\xi} + \mathbf{z}_{ij}'\mathbf{b}_i), \qquad (22)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \qquad (23)$$

$$E(\boldsymbol{\theta}_i) = E[(\theta_{i1},\ldots,\theta_{in_i})'] = \boldsymbol{\vartheta}_i, \qquad (24)$$

$$\text{Var}(\boldsymbol{\theta}_i) = \Sigma_i. \qquad (25)$$

This model has the same structure of the one by Booth et al. (2003). In the spirit of Table 1, the θ_{ij} can be assumed to follow a gamma model, producing, what we could term, a Poisson–gamma–normal model or, equivalently, a negative-binomial–normal model. When the gamma distribution is chosen, it is implicitly assumed that the components θ_{ij} of $\boldsymbol{\theta}_i$ are independent. This is natural in many cases, in the sense that the $\mathbf{b}_i$ will induce association between repeated measurements, with then the θ_{ij} taking care of additional dispersion. In this case, Σ_i reduces to a diagonal matrix. Nevertheless, it is perfectly possible to allow for general covariance structures. When a fully distributional specification would be desired, then one could choose, for example, multivariate extensions of the gamma model (Gentle, 2003).

As stated in general above, regarding the overdispersion random effects, three situations could be of interest: (1) the random-effects θ_{ij} are independent; (2) they are allowed to be dependent; (3) they are equal to each other and hence reduce to $\theta_{ij} \equiv \theta_i$.

The marginal mean vector and variance–covariance matrix are derived in Appendix B. The existence of such closed forms has important implications because they allow, for example, for explicit correlation expressions, on the one hand, and for a more versatile collection of estimation methods, on the other hand, a point to which we will return in Section 5. The availability of closed-form variance and joint-probability expressions supplements the work of, for example, Zeger, Liang and Albert (1988), who had stated that only explicit mean expressions are available for a limited number of generalized linear mixed models, other than the linear mixed model.

Let us consider strong conjugacy in this case. The corresponding model elements in Table 1 change to

$$\begin{aligned} f(\theta) &= \exp\left\{(\alpha-1)\ln\theta - \frac{1}{\beta}\theta - \ln[\beta^{\alpha}\Gamma(\alpha)]\right\},\\ f(y|\lambda=\theta\kappa) &= \exp\{y\ln\theta - \kappa\theta - \ln y! + y\ln\kappa\},\\ \phi &= 1,\\ h(\theta) &= \ln\theta,\\ g(\theta) &= \theta\kappa,\\ \gamma &= (\beta\kappa)^{-1},\\ \psi &= \beta\kappa(\alpha-1),\\ c(y,\phi) &= \ln y! + y\ln\kappa,\\ c^*(\gamma,\psi) &= (1+\psi\gamma)\ln\gamma\kappa - \ln\Gamma(1+\psi\gamma). \end{aligned}$$

Recall that the crux behind this result is (17).

Even though Molenberghs, Verbeke and Demétrio (2007) did not do so, it is fairly straightforward to derive the moments. Employing the moments' expression for the standard Poisson (Johnson, Kemp and Kotz, 2005, page 162), the expression conditional upon the random effects is

$$E(Y_{ij}^k) = \sum_{\ell=0}^{k} S(k,\ell)(\theta_{ij}\kappa_{ij})^{\ell}, \tag{26}$$

where $S(k,\ell)$ is the so-called Stirling number of the second kind. Integrating (26) over the random effects produces, without any problem,

$$\begin{aligned} E(Y_{ij}^k) = \sum_{\ell=0}^{k} S(k,\ell)\frac{\beta^{\ell}\Gamma(\alpha+\ell)}{\Gamma(\alpha)}\\ \cdot\exp\left[\ell\mathbf{x}_{ij}'\boldsymbol{\xi} + \frac{1}{2}\ell^2\mathbf{z}_{ij}'D\mathbf{z}_{ij}\right]. \end{aligned} \tag{27}$$

4.6 Specific Case: Bernoulli-Type Models for Binary Data with Logit Link

Similar to the Poisson case in Section 4.5, a natural binary-data counterpart to (21)–(25) is

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij} = \theta_{ij}\kappa_{ij}), \tag{28}$$

$$\kappa_{ij} = \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\xi} + \mathbf{z}_{ij}'\mathbf{b}_i)}{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\xi} + \mathbf{z}_{ij}'\mathbf{b}_i)}, \tag{29}$$

completing the specification with (23)–(25). Unlike in the Poisson case, closed forms for neither the mean nor the variance follow when normal random effects are present. When only overdispersion random effects are included, especially when they are assumed to follow a beta distribution, as in Table 1, conjugacy applies. However, the beta distribution does not allow for the multiplicative invariance as (17), which will preclude strong conjugacy.

When the overdispersion random effects are assumed to be equal, $\theta_{ij} = \theta_i$, then the beta–binomial model would follow if no normal random effects are present. The same is true, by the way, for the compound-symmetry model generated by the hierarchical random-intercepts model in the Gaussian case.

Explicitly considering $\theta_{ij} \sim \text{Beta}(\alpha,\beta)$, then $\phi_{ij} = \alpha/(\alpha+\beta)$, and

$$\begin{aligned} \sigma_{ij}^2 = \sigma_{i,jj} &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},\\ \sigma_{i,jk} &= \rho_{ijk}\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \end{aligned}$$

Observe that there are two correlations: ρ_{ijk}, which described the correlation between draws from the beta distribution and $(\alpha+\beta+1)^{-1}$. It is of course possible to let α and β vary with i and/or j. In such cases, the above and below expressions will change somewhat, but computations are straightforward.

Using the general expressions, the above results can be used to derive approximate expressions for means and variance–covariance elements. For the special case of no normal random effects, but maintaining the fixed effects in (29), that is,

$$\kappa_{ij} = \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\xi})}{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\xi})}, \tag{30}$$

we obtain

$$\begin{aligned} E(Y_{ij}) &= \frac{\alpha}{\alpha+\beta}\kappa_{ij},\\ \text{Var}(Y_{ij}) &= \frac{\alpha}{\alpha+\beta}\kappa_{ij} - \left(\frac{\alpha}{\alpha+\beta}\right)\kappa_{ij}^2,\\ \text{Cov}(Y_{ij},Y_{ik}) &= \rho_{ijk}\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\kappa_{ij}\kappa_{ik}. \end{aligned} \tag{31}$$

If we further make exchangeability assumptions, that is, $\kappa_{ij} = \kappa_{ik} \equiv \kappa_i$ and $\rho_{ijk} = \rho_i$, further simplification follows. Finally, setting $\kappa_i = 1$, the conventional beta-binomial follows. It is then easy to derive the resulting binomial version by defining

$$Z_i = \sum_{i=1}^{n_i} Y_{ij}. \tag{32}$$

Simple algebra then shows

$$E(Z_i) = n_i\frac{\alpha}{\alpha+\beta} = n_i\pi_i,$$

$$\begin{aligned}\text{Var}(Z_i) &= n_i \frac{\alpha\beta}{(\alpha+\beta)^2}\left\{1+(n_i+1)\frac{1}{\alpha+\beta+1}\right\}\\ &= n_i\pi_i(1-\pi_i)\{1+(n_i-1)\widetilde{\rho}_i\},\end{aligned}$$

with $\widetilde{\rho}_i$ the beta-binomial correlation. Hence, the conventional beta-binomial model follows.

In comparison to the longitudinal Poisson case, the longitudinal binary case appears to defeat closed-form solutions and strong conjugacy. However, this hinges on the fact that we employ the logit link. In spite of it being a very natural choice in the univariate case, it does not combine very nicely with normal random effects. Recall that this is known already from the GLMM framework for binary data. Therefore, it is sensible to study the probit link instead. The random-effects probit model has received some attention in earlier decades (Schall, 1991; Guilkey and Murphy, 1993; Hedeker and Gibbons, 1994; McCulloch, 1994; Gibbons and Hedeker, 1997; Renard, Molenberghs and Geys, 2004), with emphasis primarily on computational schemata to deal with the multivariate normal integral. We will return to this aspect in Section 5.

4.7 Specific Case: Bernoulli-Type Models for Binary Data with Probit Link

Introducing the probit version of the model, while at the same time assuming that the overdispersion parameters are beta distributed, comes down to

$$\kappa_{ij} = \Phi_1(\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i), \tag{33}$$

$$\theta_{ij} \sim \text{Beta}(\alpha, \beta). \tag{34}$$

Like before, α and β could be allowed to vary with i and/or j.

It now follows that the joint distribution can be written as (details in Appendix D)

$$f_{n_i}(\mathbf{y}_i = \mathbf{1}) = \left(\frac{\alpha}{\alpha+\beta}\right)^{n_i} \cdot \Phi_{n_i}(X_i\boldsymbol{\xi}; L_{n_i}^{-1}), \tag{35}$$

with

$$L_{n_i} = I_{n_i} - Z_i(D^{-1} + Z_i'Z_i)^{-1}Z_i'. \tag{36}$$

More details on the cell probabilities, as well as on means and variances, can be found in Appendix D.

It is important to note that the existence of closed-form expressions for the probit case opens a window of opportunity for the logit case. Indeed, the well-known approximation formulae, linking the normal and logistic densities, proves useful here. As shown in Johnson and Kotz (1970), page 6, and used in Zeger, Liang and Albert (1988),

$$\frac{e^y}{1+e^y} \approx \Phi_1(cy), \tag{37}$$

with $c = (16\sqrt{3})/(15\pi)$. Applied to (28)–(29), we find

$$\begin{aligned}\pi_{ij} &\sim \theta_{ij}\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)}{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)}\\ &\approx \theta_{ij}\Phi_1[c(\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)].\end{aligned} \tag{38}$$

Applying (38) to (35) yields

$$f_{n_i}(\mathbf{y}_i = \mathbf{1}) \approx \left(\frac{\alpha}{\alpha+\beta}\right)^{n_i} \cdot \Phi_{n_i}(cX_i\boldsymbol{\xi}; \widetilde{L}_{n_i}^{-1}), \tag{39}$$

with

$$\widetilde{L}_{n_i} = I_{n_i} - c^2 Z_i(D^{-1} + Z_i'Z_i)^{-1}Z_i'.$$

For the expectation, we find, based on (38) and (D.4)

$$\begin{aligned}E(Y_{ij}) &\approx \frac{\alpha}{\alpha+\beta}\\ &\quad\cdot \Phi_1(|I + c^2 D z_{ij}z_{ij}'|^{-1/2}c\mathbf{x}'_{ij}\boldsymbol{\xi}),\end{aligned} \tag{40}$$

with similar expressions for the variance and covariance terms. Note that, upon estimating the parameters within the probit approximation paradigm, back-transformation to the original logit scale is possible, using expressions such as (38) and (40). This opens perspectives for alternative estimation methods for the combined model with logit link, with the important special case of the normal-logistic GLMM.

In the Bernoulli case, calculating the moments is extremely simple. Indeed, the Bernoulli moments are all identical. The conditional moments are all $E(Y_{ij}^k|\theta_{ij}, \mathbf{b}_i) = \theta_{ij}\kappa_{ij}$ $(k = 1, 2, \ldots)$. Hence, they all reduce to (31). In the probit case, they equal to (D.4).

4.8 Specific Case: Weibull- and Exponential-Type Models for Time-to-Event Data

The general Weibull model for repeated measures, with both gamma and normal random effects, can be expressed as

$$\begin{aligned}f(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{b}_i) &= \prod_{j=1}^{n_i} \lambda\rho\theta_{ij}y_{ij}^{\rho-1}e^{\mathbf{x}'_{ij}\boldsymbol{\xi}+\mathbf{z}'_{ij}\mathbf{b}_i}\\ &\quad\cdot e^{-\lambda y_{ij}^{\rho}\theta_{ij}e^{\mathbf{x}'_{ij}\boldsymbol{\xi}+\mathbf{z}'_{ij}\mathbf{b}_i}},\end{aligned} \tag{41}$$

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)}\theta_{ij}^{\alpha_j-1}e^{-\theta_{ij}/\beta_j}, \tag{42}$$

$$f(\mathbf{b}_i) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}}e^{-(1/2)\mathbf{b}_i'D^{-1}\mathbf{b}_i}. \tag{43}$$

A few observations are in place. First, it is implicit that the gamma random effects are independent. This need

not be the case and, like in the Poisson case, extension via multivariate gamma distributions is possible. Second, setting $\rho = 1$ leads to the special case of an exponential time-to-event distribution. Third, it is evident that the classical gamma frailty model (i.e., no normal random effects) and the Weibull-based GLMM (i.e., no gamma random effects) follow as special cases. Fourth, owing to the conjugacy result of Table 1 and property (17) of the gamma density, strong conjugacy applies. This is typically considered for the exponential model, but it holds for the Weibull model too, merely by observing that the Weibull model is nothing but an exponential model for the random variable Y_{ij}^{ρ}. It is equally possible to derive this result by merely rewriting the factor $\phi = \lambda\kappa$. Fifth, the above expressions are derived for a two-parameter gamma density. It is customary in a gamma frailty context (Duchateau and Janssen, 2007) to set $\alpha_j\beta_j = 1$, for reasons of identifiability. In this case, (42) is replaced by

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{(1/\alpha_j)^{\alpha_j}\Gamma(\alpha_j)} \theta_{ij}^{\alpha_j - 1} e^{-\alpha_j \theta_{ij}}. \tag{44}$$

Alternatively, assuming $\alpha_j = 1$ and $\beta_j = 1/\delta_j$, one could write

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \delta_j e^{-\delta_j \theta_{ij}}, \tag{45}$$

implying that the gamma density is reduced to an exponential one. Closed-form expressions for the marginal density, means, variances, covariances and moments are derived in Appendix E, where also a number of related facts are derived.

Of course, in this context of time-to-event data, further issues that deserve attention are as follows: (1) censoring and how to deal with it; (2) derivation of related functions, such as the survivorship function, as well as the hazard, cumulative hazard and intensity functions; (3) the possibility of nonparametric baseline hazard functions. These are nevertheless not considered here. While in principle possible, we aim at focusing on commonality between various GLM settings.

4.9 Implication for Computation of Correlation and Derived Quantities

Up to here, we have provided closed-form expressions for the marginal joint distributions, the moments, and hence for means and variances, for the normal, Poisson, probit and Weibull cases, with a combination of normal random effects, on the one hand, supplemented, on the other hand, with conjugate random effects, taking a normal, gamma, beta and gamma form, respectively. The obvious one missing from the list is the logit model, but then the logit-probit connection, as discussed in Section 4.7, comes to rescue. Generally, progress is possible whenever strong conjugacy applies.

These results and the ensuing calculations are useful for a number of reasons, such as: (1) parameter estimation and derived inferences; (2) implementation of estimation algorithms, as will be discussed in Section 5; and (3) the computation of derived quantities.

Such derived quantities include marginal correlation coefficients, about which more detail is provided in the Appendix (Section F). Of course, correlations are not always of direct scientific interest and, when they are, one might not be willing to base one's entire model choice on whether or not closed-form correlations are available. That said, some considerations are in place.

First, our results indicate that closed-form correlations exist for a number of commonly used models, such as the Poisson–normal GLMM and the Weibull–gamma frailty model. Second, the same holds true for their extensions within our proposed model. Third, when studying psychometric reliability and generalizability (Vangeneugden et al., 2008a, 2010), the correlation function is the basic building block. Fourth, correlation functions are also used in the context of surrogate marker evaluation from clinical-trial data (Burzykowski, Molenberghs and Buyse, 2005).

At the same time, the one important situation that evades direct calculation of the marginal correlation is the logit with beta and normal random effects, but then the probit–logit correspondence can be invoked. On the one hand, the probit link can be used in lieu of the logit link; on the other hand, the calculations can be carried out on the probit scale, where after the results can be back transformed to the logit scale.

Other key derived quantities include marginal regression parameters. Suppose, for example, that one is interested in estimating the marginal treatment effect from longitudinal clinical-trial data that are not normally distributed. In principle, a marginal model could be fitted, which oftentimes is done via generalized estimating equations (Liang and Zeger, 1986). However, when data are incomplete, such models pose specific challenges even though remedies have been devised, such as inverse probability weighting or a combination with multiple imputation (for reviews, see Fitzmaurice et al., 2009). These, however, come with their own problems. It is then attractive to fit a GLMM, with or without additional random effects for overdispersion, and use the closed-form mean expressions to derive

marginal mean function. The estimand of interest is then $E(\mathbf{Y}_i|T_i=1)-E(\mathbf{Y}_i|T_i=0)$, where T_i is the obvious indicator for the treatment to which the ith subject has been assigned. Precision estimation then proceeds via the delta method.

5. ESTIMATION

A priori, fitting a combined model of the type described in Section 4 proceeds by integrating over the random effects. The likelihood contribution of subject i is

$$\begin{aligned} &f_i(\mathbf{y}_i|\boldsymbol{\vartheta}, D, \boldsymbol{\vartheta}_i, \Sigma_i) \\ &\quad = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) f(\mathbf{b}_i|D) \\ &\qquad \cdot f(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}_i, \Sigma_i)\, d\mathbf{b}_i\, d\boldsymbol{\theta}_i. \end{aligned} \tag{46}$$

Here, $\boldsymbol{\vartheta}$ groups all parameters in the conditional model for $\mathbf{Y}_i$. From (46) the likelihood derives as

$$\begin{aligned} &L(\boldsymbol{\vartheta}, D, \boldsymbol{\vartheta}, \Sigma) \\ &\quad = \prod_{i=1}^{N} f_i(\mathbf{y}_i|\boldsymbol{\vartheta}, D, \boldsymbol{\vartheta}_i, \Sigma_i) \\ &\quad = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) f(\mathbf{b}_i|D) \\ &\qquad \cdot f(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}_i, \Sigma_i)\, d\mathbf{b}_i\, d\boldsymbol{\theta}_i. \end{aligned} \tag{47}$$

The key problem in maximizing (47) is the presence of N integrals over the random effects $\mathbf{b}_i$ and $\boldsymbol{\theta}$. It is widely claimed that the absence of a closed-form solution precludes an analytical-integration based solution (Molenberghs and Verbeke, 2005), explaining the popularity of Taylor-series expansion based methods, such as PQL and MQL, Laplace approximation and numerical-integration based methods. These have been implemented in, for example, the SAS procedures GLIMMIX and NLMIXED. Several of the series expansion methods tend to exhibit bias, an issue taken up in Breslow and Lin (1995), and suggesting the use of alternative methods.

However, thanks to our results in Section 4, further progress can be made. Closed-form integration, apart from the normal case, is within reach for the Poisson, probit and Weibull cases. Now, some closed forms involve series expansions, and may be either time consuming or cumbersome to implement. This notwithstanding, a variety of alternative approaches are possible.

Let us turn to the Poisson case. While closed-form expressions can be used to implement maximum likelihood estimation, with numerical accuracy governed by the number of terms included in the series, one can also proceed by what we will term partial marginalization. By this we refer to integrating (21)–(25) over the gamma random effects only, leaving the normal random effects untouched. The corresponding probability is

$$\begin{aligned} f(y_{ij}|\mathbf{b}_i) &= \binom{\alpha_j + y_{ij} - 1}{\alpha_j - 1} \cdot \left(\frac{\beta_j}{1+\kappa_{ij}\beta_j}\right)^{y_{ij}} \\ &\quad \cdot \left(\frac{1}{1+\kappa_{ij}\beta_j}\right)^{\alpha_j} \kappa_{ij}^{y_{ij}}, \end{aligned} \tag{48}$$

where $\kappa_{ij} = \exp[\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i]$. Note that, with this approach, we assume that the gamma random effects are independent within a subject. This is fine, given the correlation is induced by the normal random effects.

Similarly, for the Weibull case we obtain

$$f(y_{ij}|\mathbf{b}_i) = \frac{\lambda\kappa_{ij}e^{\mu_{ij}}\rho y_{ij}^{\rho-1}\alpha_j\beta_j}{(1+\lambda\kappa_{ij}e^{\mu_{ij}}\beta_j y_{ij}^{\rho})^{\alpha_j+1}}. \tag{49}$$

Because there is lack of strong conjugacy, the logit case defies the mere exploitation of conjugate form, such as the negative-binomial form (48) and the Weibull–gamma frailty form (49). Nevertheless, it is easy to derive, for this case,

$$\begin{aligned} f(y_{ij}|\mathbf{b}_i) &= \frac{1}{\alpha_j+\beta_j} \cdot (\kappa_{ij}\alpha_j)^{y_{ij}} \\ &\quad \cdot [(1-\kappa_{ij})\alpha_j + \beta_j]^{1-y_{ij}}. \end{aligned} \tag{50}$$

For all of these, it is straightforward to obtain the fully marginalized probability by numerically integrating the normal random effects out of (48), (49) and (50), using a tool such as the SAS procedure NLMIXED that allows for normal random effects in arbitrary, user-specified models.

The concept of partial integration always applies whenever strong conjugacy holds. Indeed, an expression of the form (16) corresponds to integrating over the conjugate random effect θ, while leaving the normally distributed random effect embedded in the predictor, κ in this notation. Recall that, while expressions of the type (16) appear to be for the univariate case, they extend without problem to the longitudinal setting as well.

For the specific case of the marginalized probit model, the computational challenge stems from the presence of a multivariate normal integral of the

form (35), a phenomenon also known from the fully marginally specified multivariate probit model (Ashford and Sowden, 1970; Lesaffre and Molenberghs, 1991; Molenberghs and Verbeke, 2005). Specific to the context of the probit models with random effects, Zeger, Liang and Albert (1988) derived the marginal mean function, needed for their application of generalized estimating equations as a fitting algorithm for the marginalized probit model. It is one of the first instances of the use of GEE to a nonmarginally specified model. Precisely, these authors derive the marginal mean function and (a working version of) the marginal variance–covariance matrix. These are sufficient to implement GEE or, with appropriate extension, also second-order GEE. Note that our derivations yield, for strong conjugate cases in general, as well as for a number of particular cases, not only the marginal mean and variance, but also all moments and the entire joint distribution. Evidently, this is plenty to implement GEE, but the other methods, described in this section, come within reach, too.

In the same spirit, pseudo-likelihood can be used (Aerts et al., 2002; Molenberghs and Verbeke, 2005). This is particularly useful when the joint marginal distribution is available but cumbersome to manipulate and evaluate, such as in the probit case. This is the idea followed by Renard, Molenberghs and Geys (2004) for a multilevel probit model with random effects, similar in spirit to the probit models considered in Section 4.7. Essentially, the joint distribution is replaced with a product of factors of marginal and/or conditional distributions of lower dimensions. Because such a product does not necessarily recompose the original joint distribution, sandwich-estimator ideas are then used to provide not only valid point estimates, but also precision estimates and inferences derived therefrom.

Schall (1991) proposed an efficient and general estimation algorithm, based on Harville's (1974) modification of Henderson's (1984) mixed-model equations. Hedeker and Gibbons (1994) and Gibbons and Hedeker (1997) proposed numerical-integration based methods, thus considering neither marginal moments (means, variances) nor marginalized joint probabilities. Guilkey and Murphy (1993) provide a useful early overview of estimation methods and then revert to Butler and Moffit's (1982) Hermite-integration based method, supplemented with Monte Carlo Markov Chain ideas.

Further, one might, for example, opt for fully Bayesian inferences. Alternatively, the EM algorithm can be used, in line with Booth et al. (2003) for the Poisson case. The EM is a flexible framework within which either the conjugate, or the normal, or both sets of random effects can be considered the "missing" data over which expectations are taken.

Booth et al. (2003) also considered nonparametric maximum likelihood, in the spirit of Aitkin (1999) and Alfò and Aitkin (2000). In addition, ideas of hierarchical generalized linear models (Lee and Nelder, 1996, 2001a, 2001b, 2003; Yun, Sohn and Lee, 2006; Lee, Nelder and Pawitan, 2006) can be employed.

A suite of methods is available that employ transformation results, essentially based on transforming the nonnormal random effects to normal ones, or vice versa. To briefly describe these, write the contribution for subject i to the likelihood as

$$L_i = \int \left[\prod_j f(y_{ij}|u_i)\right] p_u(u_i)\, du_i, \tag{51}$$

where $f(\cdot)$ specifies the outcome model given the random effects. Furthermore, $p_u(\cdot)$ denotes the density of the random effect, typically nonnormal. While the latter random effect can be vector-valued, let us illustrate the method for the scalar case. To simplify notation further, in (51), covariates and parameter vectors have been suppressed from notation. Liu and Yu (2008) advocate a simple transformation:

$$L_i = \int \left[\prod_j f(y_{ij}|a_i)\right] \frac{p_u(a_i)}{\phi(a_i)} \phi(a_i)\, da_i, \tag{52}$$

where now a_i is a normal random effect. Evidently, $\phi(\cdot)$ is the standard (multivariate or univariate) normal density. Liu and Yu (2008) complete their argument by stating that then the new model

$$\left[\prod_j f(y_{ij}|a_i)\right] \frac{p_u(a_i)}{\phi(a_i)}$$

can be subjected to the conventional quadrature techniques available in, for example, SAS' NLMIXED procedure. A number of SAS implementations for important particular cases are offered by these authors. Obviously, the method can be expanded to our situation, where apart from the nonnormal random effects, also normal random effects are present. The justification of the method simply follows by applying the transformation theorem at the level of the densities involved. The usefulness of this method cannot be overestimated. It is especially useful when partial integration is not possible, for example, when strong conjugacy does not hold, like in the binary beta–normal–logit case.

Alternatively, Nelson et al. (2006) advocate the transformation

$$u_i = F_u^{-1}[(\Phi(a_i)], \tag{53}$$

where F_u is the cumulative distribution function (CDF) of u_i and $\Phi(\cdot)$ is the standard normal CDF, as before. Nelson et al.'s method, labeled *probability integral transformation* (PIT), comes down to generating normal variates and then inserting these in the model only after transformation (53), ensuring that they are of the desired nature. It is tautologically clear that (53) automatically ensures the support of the variable is correctly mapped along with the variable itself. By passing through the unit interval, by means of $\Phi(\cdot)$, and then applying $F_u(\cdot)$, one forces, for example, a gamma variable to range over the positive half line, a beta variable to be confined to the unit interval, etc., as it should.

Lin and Lee (2008) present estimation methods for the specific case of linear mixed models with skew-normal, rather than normal, random effects.

Quite apart from the choice of estimation method, it is important to realize that not all parameters may be simultaneously identifiable. For example, the gamma-distribution parameters in the Poisson case, α and β, are not simultaneously identifiable when the linear-predictor part is also present, because there is aliasing with the intercept term. Therefore, one can set, for example, β equal to a constant, removing the identifiability problem. It is then clear that α, in the univariate case, or the set of α_j in the repeated-measures case, describe the additional overdispersion, in addition to what stems from the normal random effect(s). A similar phenomenon also plays in the binary case, where both beta-distribution parameters are not simultaneously estimable.

6. ANALYSIS OF CASE STUDIES

6.1 A Clinical Trial in Epileptic Patients

We will analyze the epilepsy data, introduced in Section 2.1. Note that the data were analyzed before in Molenberghs and Verbeke (2005), Chapter 19, using generalized estimating equations (Liang and Zeger, 1986) and the Poisson–normal model. These authors used a slightly different parameterization.

Let Y_{ij} represent the number of epileptic seizures patient i experiences during week j of the follow-up period. Also, let t_{ij} be the time-point at which Y_{ij} has been measured, $t_{ij} = 1, 2, \ldots,$ until at most 27. Let us consider the combined model (21)–(25), with specific choices

$$\ln(\kappa_{ij}) = \begin{cases} (\xi_{00} + b_i) + \xi_{01} t_{ij}, & \text{if placebo,} \\ (\xi_{10} + b_i) + \xi_{11} t_{ij}, & \text{if treated,} \end{cases} \tag{54}$$

where the random intercept b_i is assumed to be zero-mean normally distributed with variance d. We consider special cases: (1) the ordinary Poisson model, (2) the negative-binomial model, (3) the Poisson–normal model, together with (4) the combined model. Estimates (standard errors) are presented in Table 2. Clearly, both the negative-binomial model and the Poisson–normal model are important improvements, in terms of the likelihood, relative to the ordinary Poisson model. This should come as no surprise since the latter unrealistically assumes there is neither overdispersion nor correlation within the outcomes, while clearly both are present. In addition, when considering the combined model, there is a very strong improvement in fit when gamma and normal random effects are simultaneously allowed for. This strongly affects the point and precision estimates of such key parameters as the slope difference and the slope ratio. There is also an impact on hypothesis testing. The Poisson model leads to unequivocal significance for both the difference ($p = 0.0008$) and ratio ($p = 0.0038$), whereas for the Poisson–normal this is not the case for the difference of the slopes ($p = 0.7115$), while some significance is maintained for the ratio ($p = 0.0376$). Because the Poisson–normal is commonly used, it is likely that in practice one would decide in favor of a treatment effect when considering the slope ratio. This is no longer true with the negative-binomial model, where the p-values change to $p = 0.01310$ and $p = 0.2815$, respectively. Of course, one must not forget that, while the negative-binomial model accommodates overdispersion, the θ_{ij} random effects are assumed independent, implying independence between repeated measures. Again, this is not realistic and, therefore, the combined model is a more viable candidate, corroborated further by the aforementioned likelihood comparison. This model produces nonsignificant p-values of $p = 0.2260$ and $p = 0.1591$, respectively.

Thus, in conclusion, whereas the conventionally used and broadly implemented Poisson–normal model would suggest a significant effect of treatment, our combined model issues a message of caution, because there is no evidence whatsoever regarding a treatment difference.

TABLE 2

Epilepsy study. Parameter estimates and standard errors for the regression coefficients in (1) *the Poisson model*, (2) *the negative-binomial model*, (3) *the Poisson–normal model and* (4) *the combined model. Estimation was done by maximum likelihood using numerical integration over the normal random effect, if present*

		Estimate (s.e.)	
Effect	**Parameter**	**Poisson**	**Negative-binomial**
Intercept placebo	ξ_{00}	1.2662 (0.0424)	1.2594 (0.1119)
Slope placebo	ξ_{01}	−0.0134 (0.0043)	−0.0126 (0.0111)
Intercept treatment	ξ_{10}	1.4531 (0.0383)	1.4750 (0.1093)
Slope treatment	ξ_{11}	−0.0328 (0.0038)	−0.0352 (0.0101)
Negative-binomial parameter	α_1	—	0.5274 (0.0255)
Negative-binomial parameter	$\alpha_2 = 1/\alpha_1$	—	1.8961 (0.0918)
−2log-likelihood		−1492	−6755
		Poisson–normal	Combined
Intercept placebo	ξ_0	0.8179 (0.1677)	0.9112 (0.1755)
Slope placebo	ξ_1	−0.0143 (0.0044)	−0.0248 (0.0077)
Intercept treatment	ξ_0	0.6475 (0.1701)	0.6555 (0.1782)
Slope treatment	ξ_2	−0.0120 (0.0043)	−0.0118 (0.0074)
Negative-binomial parameter	α_1	—	2.4640 (0.2113)
Negative-binomial parameter	$\alpha_2 = 1/\alpha_1$	—	0.4059 (0.0348)
Variance of random intercepts	d	1.1568 (0.1844)	1.1289 (0.1850)
−2log-likelihood		−6810	−7664

Molenberghs and Verbeke (2005), Chapter 19, considered a Poisson–normal model with random intercepts as well as random slopes in time. It is interesting to note that, when allowing for such an extension in our models, the random slopes improve the fit of the Poisson–normal model with random intercept, but not of the combined one with random intercept (details not shown). As a consequence, the combined model with random intercept is the best fitting one. At the same time, note that fitting such a model establishes that the presence of a conjugate random effect does not preclude the consideration of normal random effects beyond random intercepts.

Recall that the data were analyzed, too, by Booth et al. (2003). While we considered four different models, these authors focused on the Poisson–normal and combined implementations. There are further differences in actual fixed-effects and random-effects models considered, as well as in us further considering inferences for differences and ratios.

Let us now turn to the correlation functions. Given that the gamma random effects are assumed independent, we only need to consider the Poisson–normal and combined cases; the versions with and without random slopes are considered. Obviously, because the fixed-effects structure is not constant but rather depends on time, we have to apply the general correlation function (F.13). In the Poisson–normal case with random intercepts only, and for the placebo group, based on the parameter estimates in Table 2, we obtain

$$\begin{aligned}\text{Corr}(Y(t), Y(s)) = {} & 35.58 \cdot 0.99^{t+s} \\ & /\big(\sqrt{(4.04 \cdot 0.99^t + 35.58 \cdot 0.97^t)} \\ & \cdot \sqrt{(4.04 \cdot 0.99^s + 35.58 \cdot 0.97^s)}\big),\end{aligned}$$

where $Y(t)$ represents the outcome for an arbitrary subject at time t. Calculations in all other cases are similar. The smallest and largest values for the correlation functions, for both arms, for both the Poisson–normal and combined models, and for both choices of the random-effects structure are given in Table 3. When only random intercepts are considered, the correlations range over a narrow interval; they are rather high and there is little difference between the Poisson–normal and combined models. However, turning to the models with random intercepts and random slopes, several differences become apparent. First, the values exhibit a much broader range between their smallest and largest values. Second, the range is somewhat overestimated by the Poisson–normal model, which then narrows when we switch to the combined model, thereby incorporating overdispersion effects, random intercepts and ran-

TABLE 3
Epilepsy study. Observed smallest and largest values for the correlation function, for the Poisson–normal and combined models, and for both treatment arms. The time pair for which the values are observed is shown too (RI—random intercept; RS—random slope)

Model	Arm	Smallest value		Largest value	
		ρ	Time pair	ρ	Time pair
Poisson–normal, RI	Placebo	0.8577	26 & 27	0.8960	1 & 2
Poisson–normal, RI	Treatment	0.8438	26 & 27	0.8794	1 & 2
Combined, RI	Placebo	0.8259	26 & 27	0.8981	1 & 2
Combined, RI	Treatment	0.8383	26 & 27	0.8744	1 & 2
Poisson–normal, RI+RS	Placebo	0.2966	1 & 27	0.9512	26 & 27
Poisson–normal, RI+RS	Treatment	0.2936	1 & 27	0.9530	26 & 27
Combined, RI+RS	Placebo	0.4268	1 & 27	0.9281	26 & 27
Combined, RI+RS	Treatment	0.4225	1 & 27	0.9329	26 & 27

dom slopes. Thus, the random slope allows for the correlation to range over a considerable interval, while the overdispersion effect avoids the range to be overly wide.

Within each model, there is relatively little difference between the placebo and treated groups, although the difference is a bit more pronounced in the combined model. Further, the correlation range within every group is relatively narrow. The most noteworthy feature, unquestionably, is the large discrepancy between both models. This is because the Poisson–normal model forces the correlation and overdispersion effects to stem from a single additional parameter, the random-intercept variance d. Thus, considerable overdispersion also forces the correlation to increase, arguably beyond what is consistent with the data. In the combined model, in contrast, there are *two* additional parameters, giving proper justice to both correlation and overdispersion effects. It was already clear from the above discussion and that in Molenberghs, Verbeke and Demétrio (2007) that the combined model is an important improvement. This now clearly manifests itself in the correlation function, too.

6.2 A Clinical Trial in Onychomycosis

We will analyze the binary onychomycosis data, introduced in Section 2.2. For the logit, consider the model

$$\begin{aligned} Y_{ij}|(b_i) &\sim \text{Bernoulli}(\pi_{ij}),\\ \text{logit}(\pi_{ij}) &= \xi_1(1-T_i) + b_i + \xi_2(1-T_i)t_{ij}\\ &\quad + \xi_3 T_i + \xi_4 T_i t_{ij}, \end{aligned} \tag{55}$$

where T_i is the treatment indicator for subject i, t_{ij} is the time-point at which the jth measurement is taken for the ith subject, and $b_i \sim N(0,d)$. Parameter estimates for the logistic model, with and without the normal random effect, on the one hand, and with and without the beta–binomial component, on the other hand, as described in Section 4.6, are presented in Table 4. Observe that the model becomes hard to fit when the beta random effects are present, which is seen from estimates and standard errors in both the beta–binomial model as well as the combined model. To understand this, we must observe that the conjugate random effects in the Bernoulli case, unlike in the Poisson, binomial and Weibull cases, cannot add to the variability, only to the correlation structure. This means that there is considerably less information available than in the other cases. This does not mean that the beta random effects are unnecessary, but rather that they challenge the stable estimation of other model parameters.

6.3 Recurrent Asthma Attacks in Children

We will analyze the times-to-event, introduced in Section 2.3. We consider an exponential model, that is, a model of the form (41) with $\rho = 1$, and further a predictor of the form

$$\kappa_{ij} = \xi_0 + b_i + \xi_1 T_i,$$

where T_i is an indicator for treatment and $b_i \sim N(0,d)$. Results from fitting all four models (with/without normal random effect; with/without gamma random effect) can be found in Table 5. A formal assessment of the treatment effect from all four models is given in Table 6. The treatment effect ξ_1 is stably

TABLE 4

Onychomycosis study. Parameter estimates and standard errors for the regression coefficients in (1) *the logistic model*, (2) *the beta–binomial model*, (3) *the logistic–normal model and* (4) *the combined model. Estimation was done by maximum likelihood using numerical integration over the normal random effect, if present*

		Estimate (s.e.)	
Effect	**Parameter**	**Logistic**	**Beta–binomial**
Intercept treatment A	ξ_0	−0.5571 (0.1090)	17.9714 (1482.6)
Slope treatment A	ξ_1	−0.1769 (0.0246)	5.2454 (12970.0)
Intercept treatment B	ξ_2	−0.5335 (0.1122)	18.6744 (2077.13)
Slope treatment B	ξ_3	−0.2549 (0.0309)	4.7775 (12912.0)
Std. dev. random effect	$\sqrt{d}$	—	—
Ratio	α/β	—	3.6739 (0.2051)
−2log-likelihood		1812	1980
		Logistic–normal	Combined
Intercept treatment A	ξ_0	−1.6299 (0.4354)	−1.6042 (4.0263)
Slope treatment A	ξ_1	−0.4042 (0.0460)	−6.4783 (1.4386)
Intercept treatment B	ξ_2	−1.7486 (0.4478)	−16.2079 (3.5830)
Slope treatment B	ξ_3	−0.5634 (0.0602)	−8.0745 (1.5997)
Std. dev. random effect	$\sqrt{d}$	4.0150 (0.3812)	60.8835 (14.2237)
Ratio	α/β	—	0.2805 (0.0350)
−2log-likelihood		1248	1240

identifiable in all four models. As can be seen from Table 6, the treatment effects are similar in strengths, but including both random effects reduces the evidence, relative to the exponential model. Needless to say, too parsimonious an association structure might lead to liberal test behavior.

6.4 The Need for the Combined Model

We have some evidence from the above three examples that there is a need for the combined model. Some indication came, for example, from the correlation functions in the epilepsy case. It is useful to per-

TABLE 5

Asthma study. Parameter estimates and standard errors for the regression coefficients in (1) *the exponential model*, (2) *the exponential–gamma model*, (3) *the exponential–normal model and* (4) *the combined model. Estimation was done by maximum likelihood using numerical integration over the normal random effect, if present*

		Estimate (s.e.)	
Effect	**Parameter**	**Exponential**	**Exponential–gamma**
Intercept	ξ_0	−3.3709 (0.0772)	−3.9782 (15.354)
Treatment effect	ξ_1	−0.0726 (0.0475)	−0.0755 (0.0605)
Shape parameter	λ	0.8140 (0.0149)	1.0490 (16.106)
Std. dev. random effect	$\sqrt{d}$	—	—
Gamma parameter	γ	—	3.3192 (0.3885)
−2log-likelihood		18,693	18,715
		Exponential–normal	Combined
Intercept	ξ_0	−3.8095 (0.1028)	3.9923 (20.337)
Treatment effect	ξ_1	−0.0825 (0.0731)	−0.0887 (0.0842)
Shape parameter	λ	0.8882 (0.0180)	0.8130 (16.535)
Std. dev. random effect	$\sqrt{d}$	0.4097 (0.0386)	0.4720 (0.0416)
Gamma parameter	γ	—	6.8414 (1.7146)
−2log-likelihood		18,611	18,629

TABLE 6
Asthma study. Wald test results for the assessment of treatment effect

Model	Z-value	p-value
Exponential	−1.5283	0.1264
Exponential–gamma	−1.1293	0.2588
Exponential–normal	−1.2480	0.2120
Combined	−1.0534	0.2921

form formal comparison of all nested models, using Wald statistics, for each of the three cases. A summary is given in Table 7. Note that, owing to the familiar boundary problem that occurs when testing for variance components, mixtures of a χ_0^2 and χ_1^2 were used, instead of the conventional χ_1^2 (Molenberghs and Verbeke, 2007). In all three case studies it is clear that: (1) independence is strongly rejected in favor of both a model with normal random effects or a model with conjugate random effects; (2) on top of one set of random effects, there is a clear need for the other set as well, hence providing very strong evidence for the proposed combined model. The evidence is extremely convincing in all three cases.

These findings, taken together, imply that the data exhibit, at the same time, within-subject correlation and overdispersion.

TABLE 7
All three case studies. Wald test results for comparison of nested models

Null model	Alternative model	Z-value	p-value
Epilepsy study			
Poisson	Negative-binomial	20.68	<0.0001
Poisson	Poisson–normal	6.27	<0.0001
Negative-binomial	Combined	6.10	<0.0001
Poisson–normal	Combined	11.66	<0.0001
Onychomycosis study			
Logistic	Beta–binomial	17.91	<0.0001
Logistic	Logistic–normal	10.53	<0.0001
Beta–binomial	Combined	4.28	<0.0001
Logistic–normal	Combined	8.01	<0.0001
Asthma study			
Exponential	Exponential–gamma	8.54	<0.0001
Exponential	Exponential–normal	10.63	<0.0001
Exponential–gamma	Combined	8.54	<0.0001
Exponential–normal	Combined	3.99	<0.0001

7. CONCLUDING REMARKS

In this paper we have argued that, rather than choosing between normal and nonnormal random effects, the latter often of a gamma, beta or other conjugate type, both can usefully be integrated together into a single model, which we have termed the combined model. Our work builds upon that of Molenberghs, Verbeke and Demétrio (2007), who brought together normal random effects to induce association between repeated Poisson data, and a gamma distributed random factor in the log-linear predictor to fine tune the overdispersion. Their model produces the standard negative-binomial and Poisson–normal models as special cases, both when there are repeated measures as well as with univariate outcomes.

The current paper builds upon this work, not only by considering other important cases, such as binary and time-to-event data and, for completeness, also the normally distributed case, but, in particular, by providing an encompassing framework around it. Wherever possible, explicit expressions for the marginal joint distributions are derived, as well as for marginal means, variances, covariances and moments in general. This is possible in all cases, including the Poisson and Weibull cases, but for binary data the logit links defies such a closed form. However, we showed that a switch to the probit link does allow for closed forms. The existence of these closed forms, producing expressions for a variety of generalized linear mixed models as special cases, has not been known to its fullest extent. We discuss their implications for: (1) general understanding; (2) derived quantities such as correlations, treatment effects, etc.; and (3) the construction of parameter estimation and implementation.

For the binary case, we have exploited the logit-probit relationship to derive probit-based closed-form approximations to the logit case. For the Weibull situation, we have additionally generated a family of distributions that encompass an entire collection of Cauchy-type distributions.

To make these developments possible in their fullest generality, we have introduced *strong conjugacy*, which comes down to a version of the well-known conjugacy that is compatible with the additional introduction of normal random effects.

In terms of estimation, we have focused on maximum likelihood estimation. This can be done by integrating over the random effects, either fully analytically, using the explicit expressions derived, or by combining analytic and numeric techniques. The latter has

been implemented in the SAS procedure NLMIXED, for the Poisson, binary and survival cases, and applied to three case studies.

Of course, with the considerations of not only one but multiple sets of random effects comes the obligation to reflect on the precise nature of such latent structures. As underscored by Verbeke and Molenberghs (2009), full verification of the adequacy of a random-effects structure is not possible based on statistical considerations alone, because there is a many-to-one map from hierarchical models to the implied marginal model. Of course, this should not stop the user from considering such models, but rather issues a word of caution.

A number of topics have been mentioned in this paper that deserve further research. These include, but are not limited to, the following: (1) the construction of model building and goodness-of-fit tool; (2) a detailed study of the relative merits of various estimation methods and their implementation; (3) a study of the identifiability of (random-effects) parameters in the combined model; (4) the incorporation of censoring in the survival case; and (5) the explicit consideration of data types and models not considered here.

The Poisson, binary and Weibull cases have been implemented in the SAS procedure NLMIXED. All datasets, programs and outputs can be found in a WinZip archive on the web site www.censtat.be/software.

ACKNOWLEDGMENTS

Financial support from the IAP research network #P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. This work was partially supported by a grant from Coordenadoria para o Aperfeiçoamento de Pessoal de Nvel Superior (CAPES), Brazilian science funding agency.

SUPPLEMENTARY MATERIAL

A family of generalized linear models for repeated measures with normal and conjugate random effects: Calculation details (DOI: 10.1214/10-STS328SUPP; .pdf). In Section A, generic approximate calculations are provided. Closed-form calculations for various cases are offered as well: for the Poisson case (Section B), for the binary case with logit link (Section C), for the binary case with probit link (Section D), and for the time-to-event case (Section E). Finally, Section F is dedicated to the derivation of marginal correlation functions.

REFERENCES

AERTS, M., GEYS, H., MOLENBERGHS, G. and RYAN, L. (2002). *Topics in Modelling of Clustered Data*. Chapman & Hall, London. MR2022882

AGRESTI, A. (2002). *Categorical Data Analysis*, 2nd ed. Wiley, New York. MR1914507

AITKIN, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55** 117–128. MR1705676

ALFÒ, M. and AITKIN, M. (2000). Random coefficient models for binary longitudinal responses with attrition. *Statist. Comput.* **10** 279–288.

ASHFORD, J. R. and SOWDEN, R. R. (1970). Multivariate probit analysis. *Biometrics* **26** 535–546.

BAHADUR, R. R. (1961). A representation of the joint distribution of responses to *n* dichotomous items. In *Studies in Item Analysis and Prediction* (H. Solomon, ed.) 158–168. Stanford Univ. Press, Stanford, CA. MR0121893

BÖHNING, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others*. Chapman & Hall/CRC, London. MR1684363

BOOTH, J. G., CASELLA, G., FRIEDL, H. and HOBERT, J. P. (2003). Negative binomial loglinear mixed models. *Stat. Model.* **3** 179–181. MR2005472

BRESLOW, N. (1984). Extra-Poisson variation in log-linear models. *Appl. Statist.* **33** 38–44.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

BRESLOW, N. E. and LIN, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* **82** 81–91. MR1332840

BURZYKOWSKI, T., MOLENBERGHS, G. and BUYSE, M. (2005). *The Evaluation of Surrogate Endpoints.* Springer, New York.

BUTLER, J. S. and MOFFIT, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50** 761–765.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall/CRC, London. MR0370837

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

DEAN, C. B. (1991). Estimating equations for mixed-Poisson models. In *Estimating Functions* (V. P. Godambe, ed.) 35–46. Oxford Univ. Press, Oxford. MR1163995

DE BACKER, M., DE KEYSER, P., DE VROEY, C. and LESAFFRE, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: Terbinafine 250 mg/day vs. itraconazole 200 mg/day—a double-blind comparative trial. *British J. Dermatol.* **134** 16–17.

DUCHATEAU, L. and JANSSEN, P. (2007). *The Frailty Model*. Springer, New York.

ENGEL, B. and KEEN, A. (1994). A simple approach for the analysis of generalized linear mixed models. *Statist. Neerlandica* **48** 1–22. MR1267053

FAHRMEIR, L. and TUTZ, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. Springer, New York. MR1832899

FAUGHT, E., WILDER, B. J., RAMSAY, R. E., REIFE, R. A., KRAMER, L. D., PLEDGER, G. W. and KARIM, R. M. (1996).

Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. *Neurology* **46** 1684–1690.

FITZMAURICE, G., DAVIDIAN, M., MOLENBERGHS, G. and VERBEKE, G. (2009). *Longitudinal Data Analysis. Handbooks of Modern Statistical Methods*. Chapman & Hall/CRC, New York. MR1500110

GENTLE, J. E. (2003). *Random Number Generation and Monte Carlo Methods.* Springer, New York. MR2151519

GIBBONS, R. D. and HEDEKER, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics* **53** 1527–1537.

GUILKEY, D. K. and MURPHY, J. L. (1993). Estimation and testing in the random effects probit model. *J. Econometrics* **59** 301–317.

HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61** 383–385. MR0368279

HEDEKER, D. and GIBBONS, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **51** 933–944.

HENDERSON, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph, Canada.

HINDE, J. and DEMÉTRIO, C. G. B. (1998a). Overdispersion: Models and estimation. *Comput. Statist. Data Anal.* **27** 151–170.

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

JOHNSON, N. L., KEMP, A. and KOTZ, S. (2005). *Univariate Discrete Distributions*, 3rd ed. Wiley, Hoboken. MR2163227

JOHNSON, N. L. and KOTZ, S. (1970). *Distributions in Statistics, Continuous Univariate Distributions, Vol. 2.* Houghton-Mifflin, Boston.

KLEINMAN, J. (1973). Proportions with extraneous variance: Single and independent samples. *J. Amer. Statist. Assoc.* **68** 46–54.

LAWLESS, J. (1987). Negative binomial and mixed Poisson regression. *Canadian J. Statist.* **15** 209–225. MR0926553

LEE, Y. and NELDER, J. A. (1996). Hierarchical generalized linear models (with discussion). *J. Roy. Statist. Soc. Ser. B* **58** 619–678. MR1410182

LEE, Y. and NELDER, J. A. (2001a). Two ways of modelling overdispersion. *Appl. Statist.* **49** 591–598. MR1824561

LEE, Y. and NELDER, J. A. (2001b). Hierarchical generalized linear models: A synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika* **88** 987–1006. MR1872215

LEE, Y. and NELDER, J. A. (2003). Extended-REML estimators. *J. Appl. Statist.* **30** 845–856. MR2002126

LEE, Y., NELDER, J. A. and PAWITAN, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood.* Chapman & Hall/CRC, Boca Raton, FL. MR2259540

LESAFFRE, E. and MOLENBERGHS, G. (1991). Multivariate probit analysis: A neglected procedure in medical statistics. *Statist. Med.* **10** 1391–1403.

LIANG, K.-Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LIN, T. I. and LEE, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statist. Med.* **27** 1490–1507. MR2420253

LIU, L. and YU, Z. (2008). A likelihood reformulation method in non-normal random-effects models. *Statist. Med.* **27** 3105–3124. MR2522152

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC, London. MR0727836

MCCULLOCH, C. E. (1994). Maximum likelihood variance components estimation for binary data. *J. Amer. Statist. Assoc.* **89** 330–335.

MCLACHLAN, G. and PEEL, D. A. (2000). *Finite Mixture Models*. Wiley, New York. MR1789474

MOLENBERGHS, G. and LESAFFRE, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *J. Amer. Statist. Assoc.* **89** 633–644.

MOLENBERGHS, G. and VERBEKE, G. (2005). *Models for Discrete Longitudinal Data.* Springer, New York. MR2171048

MOLENBERGHS, G. and VERBEKE, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *Amer. Statist.* **61** 1–6. MR2339143

MOLENBERGHS, G., VERBEKE, G. and DEMÉTRIO, C. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Anal.* **13** 513–531. MR2416536

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

NELSON, K. P., LIPSITZ, S. R., FITZMAURICE, G. M., IBRAHIM, J., PARZEN, M. and STRAWDERMAN, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with non-normal random effects. *J. Comput. Graph. Statist.* **15** 39–57. MR2269362

RENARD, D., MOLENBERGHS, G. and GEYS, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Comput. Statist. Data Anal.* **44** 649–667. MR2026438

ROBERTS, D. T. (1992). Prevalence of dermatophyte onychomycosis in the United Kingdom: Results of an omnibus survey. *British J. Dermatol.* **126** (Suppl. 39) 23–27.

RIDOUT, M., DEMÉTRIO, C. G. B. and HINDE, J. (1998). Models for count data with many zeros. In *International Biometric Conference XIX* 179–192. Cape Town. Invited papers.

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

SKELLAM, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *J. Roy. Statist. Soc. Ser. B* **10** 257–261. MR0028539

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling.* Chapman & Hall/CRC, London. MR2059021

THALL, P. F. and VAIL, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46** 657–671. MR1085814

VANGENEUGDEN, T., MOLENBERGHS, G., LAENEN, A., ALONSO, A. and GEYS, H. (2008a). Generalizability in non-Gaussian longitudinal clinical trial data based on generalized linear mixed models. *J. Biopharm. Statist.* **18** 691–712. MR2523881

VANGENEUGDEN, T., MOLENBERGHS, G., VERBEKE, G. and DEMÉTRIO, C. (2010). Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Comm. Statist. Theory Methods*. To appear.

VERBEKE, G. and MOLENBERGHS, G. (2000). *Linear Mixed Models for Longitudinal Data.* Springer, New York. MR1880596

VERBEKE, G. and MOLENBERGHS, G. (2009). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statist. Model.* **00** 000–000.

WOLFINGER, R. and O'CONNELL, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *J. Statist. Comput. Simul.* **48** 233–243.

YUN, S., SOHN, S. Y. and LEE, Y. (2006). Modelling and estimating heavy-tailed non-homogeneous correlated queues Pareto-inverse gamma HGLMs with covariates. *J. Appl. Statist.* **33** 417–425. MR2224222

ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. MR0980999