Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2015, Vol. 24(4) 434–452

DOI: 10.1177/0962280214520730
smm.sagepub.com
SAGE

A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data

Geert Molenberghs,[1,2] Geert Verbeke,[1,2] Achmad Efendi,[1,3] Roel Braekers[1,2] and Clarice GB Demétrio[4]

Abstract

This paper presents, extends, and studies a model for repeated, overdispersed time-to-event outcomes, subject to censoring. Building upon work by Molenberghs, Verbeke, and Demétrio (2007) and Molenberghs et al. (2010), gamma and normal random effects are included in a Weibull model, to account for overdispersion and between-subject effects, respectively. Unlike these authors, censoring is allowed for, and two estimation methods are presented. The partial marginalization approach to full maximum likelihood of Molenberghs et al. (2010) is contrasted with pseudo-likelihood estimation. A limited simulation study is conducted to examine the relative merits of these estimation methods. The modeling framework is employed to analyze data on recurrent asthma attacks in children on the one hand and on survival in cancer patients on the other.

Keywords

exponential model, generalized Cauchy distribution, conjugacy, maximum likelihood, frailty model, pseudo-likelihood, strong conjugacy, Weibull model

1 Introduction

Time-to-event data are prominent in contemporary statistical analysis, not only for univariate outcomes but also in hierarchical settings. Apart from the need to accommodate such data hierarchies for repeated survival outcomes, recurrent events, and the like,[1] it is possible that overdispersion[2] is present in the data, relative to the standard generalized linear model[3,4] assumed, as well as censored observations.

[1]I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium
[2]I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium
[3]Study Program of Statistics, Universitas Brawijaya, Malang, Indonesia
[4]ESALQ, Piracicaba, São Paulo, Brazil

Corresponding author:
Geert Molenberghs, I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium.
Email: geert.molenberghs@uhasselt.be

While each of these features has received attention, it is uncommon to treat all of them simultaneously. Building upon their earlier work,[5] Molenberghset al.[6] presented a general modeling framework for (non-)Gaussian overdispersed and hierarchical outcomes. The time-to-event case is but one of the applications of their framework. They combine so-called conjugate random effects for overdispersion with generalized linear mixed model ideas (GLMM[7–9]) for between-subject effects. Here, we supplement their method with the possibility to accommodate censorship.

Whereas Molenberghs et al.[5] focused on maximum likelihood, using so-called partial marginalization, we supplement this inferential option with pairwise likelihood ideas.[10] A simulation study is conducted to study the relative merits of these methods. The methodology is applied to analyze data on recurrent asthma attacks in children on the one hand and on survival in cancer patients on the other.

The paper is organized as follows. In Section 2, motivating case studies with a time-to-event outcome are described, with analyses reported in Section 7. 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, with particular emphasis on so-called partial marginalization and pseudo-likelihood estimation. Some cautionary remarks regarding the existence of the corresponding marginal distributions' moments are issues in Section 6. A simulation study is described and results presented in Section 8.

2 Case studies

2.1 Recurrent asthma attacks in children

These data have been studied in Duchateau and Janssen.[1] Asthma is occurring more and more frequently in very young children (between 6 and 24 months). Therefore, a new application of an existing antiallergic 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. Data for the first two patients are listed in Table 1.

2.2 Survival in cancer patients

Hand et al.[11] presented data on patients with advanced cancer of the stomach, bronchus, colon, ovary, or breast, who were treated, in addition to standard treatment, with ascorbate. The outcome of interest, survival time in days, is recorded to address the question as to whether survival times differ with the organ being affected. Individual-patient data are listed in Table 2. There are no censored observations in this case.

Table 1. Asthma data for the first two patients.

Patient ID	Drug	Begin	End	Status
1	0	0	15	1
1	0	22	90	1
1	0	96	325	1
1	0	329	332	1
1	0	338	369	1
1	0	370	412	1
1	0	418	422	1
1	0	426	474	1
1	0	477	526	1
1	0	530	600	0
2	1	0	180	1
2	1	189	267	1
2	1	273	581	1
2	1	582	600	0

The column labeled "Status" referred to whether (1) or not (0) censoring has occurred.

Table 2. Advanced cancer data.

Stomach	Bronchus	Colon	Ovary	Breast
124	81	248	1234	1235
42	461	377	89	24
25	20	189	201	1581
45	450	1843	356	1166
412	246	180	2970	40
51	166	537	456	727
1112	63	519		3808
46	64	455		791
103	155	406		1804
876	859	365		3460
146	151	942		719
340	166	776		
396	37	372		
	223	163		
	138	101		
	72	20		
	245	283		
Average				
286.0	211.6	457.4	884.3	1395.9

Survival time in days per patient and per organ affected.

3 Background

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

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

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

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 is well known that

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

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

implying a mean–variance relationship: $\sigma^2 = \phi\psi''[\psi'^{-1}(\mu)] = \phi v(\mu)$, with $v(\cdot)$ the so-called variance function. In the exponential case, one assumes

$$f(y) = \varphi e^{-\varphi y} \tag{4}$$

with mean φ^{-1} and variance φ^{-2}. This extends in the Weibull case to

$$f(y) = \varphi\rho y^{\rho-1} e^{-\varphi y^{\rho}}$$

$$E(Y) = \varphi^{-1/\rho}\Gamma(\rho^{-1} + 1)$$

$$\mathrm{Var}(Y) = \varphi^{-2/\rho}\left[\Gamma(2\rho^{-1} + 1) - \Gamma(\rho^{-1} + 1)^2\right]$$

Note 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, $\Gamma(\cdot)$ represents the gamma function.

When the standard exponential-family models constrain the mean–variance relationship, so-called overdispersion is introduced. Early reviews are provided by Hinde and Demétrio[2] provide general treatments of overdispersion. The Poisson case received particular attention by Breslow[12] and Lawless.[13] A natural step is to allow the overdispersion parameter $\phi \neq 1$, so that (3) produces $\mathrm{Var}(Y) = \phi v(\mu)$. A convenient route is through a two-stage approach. Generally, the two-stage approach is made up of considering a distribution for the outcome, given a random effect $f(y_i|\theta_i)$ which, combined with a model for the random effect, $f(\theta_i)$, produces the marginal model

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

In our exponential and Weibull cases, it is in line with the data range to assume such a random effect to follow a gamma distribution, giving rise to the exponential-gamma and Weibull-gamma models. The model elements are listed in Table 3.

The choice of the gamma distribution can also be motivated through the concept of conjugacy.[14,15] To simplify notation, drop the indices for the purpose of the definition. The hierarchical and random-effects densities are said to be conjugate if and only if they can be written in the generic forms

$$f(y|\theta) = \exp\{\phi^{-1}[yh(\theta) - g(\theta)] + c(y, \phi)\} \tag{6}$$

$$f(\theta) = \exp\{\gamma[\psi h(\theta) - g(\theta)] + c^*(\gamma, \psi)\} \tag{7}$$

Table 3. Model elements for the exponential-gamma and Weibull-gamma models.

Element	Notation	Exponential-gamma	Weibull-gamma
Hier. model	$f(y\|\theta)$	$\varphi\theta e^{-\varphi\theta y}$	$\varphi\theta\rho y^{\rho-1}e^{-\varphi\theta y^{\rho}}$
RE model	$f(\theta)$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$
Marginal model	$f(y)$	$\frac{\varphi\alpha\beta}{(1+\varphi\beta y)^{\alpha+1}}$	$\frac{\varphi\rho y^{\rho-1}\alpha\beta}{(1+\varphi\beta y^{\rho})^{\alpha+1}}$
	$h(\theta)$	$-\theta$	$-\theta$
	$g(\theta)$	$-\ln(\theta)/\varphi$	$-\ln(\theta)/\varphi$
	ϕ	$1/\varphi$	$1/\varphi$
	γ	$\varphi(\alpha-1)$	$\varphi(\alpha-1)$
	ψ	$[\beta\varphi(\alpha-1)]^{-1}$	$[\beta\varphi(\alpha-1)]^{-1}$
	$c(y,\phi)$	$\ln(\varphi)$	$\ln(\varphi\rho y^{\rho-1})$
	$c^*(\gamma,\psi)$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi)-\ln\Gamma\left(\frac{\gamma+\varphi}{\varphi}\right)$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi)-\ln\Gamma\left(\frac{\gamma+\varphi}{\varphi}\right)$
Mean	$E(Y)$	$[\varphi(\alpha-1)\beta]^{-1}$	$\frac{\Gamma(\alpha-\rho^{-1})\Gamma(\rho^{-1}+1)}{(\varphi\beta)^{1/\rho}\Gamma(\alpha)}$
Variance	$\mathrm{Var}(Y)$	$\alpha[\varphi^2(\alpha-1)^2(\alpha-2)\beta^2]^{-1}$	$\frac{1}{\rho(\varphi\beta)^{2/\rho}\Gamma(\alpha)}\left[2\Gamma(\alpha-2\rho^{-1})\Gamma(2\rho^{-1})-\frac{\Gamma(\alpha-\rho^{-1})^2\Gamma(\rho^{-1})^2}{\rho\Gamma(\alpha)}\right]$

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 that the marginal model resulting from (6) and (7) equals

$$f(y)=\exp\left[c(y,\phi)+c^*(\gamma,\psi)-c^*\left(\phi^{-1}+\gamma,\frac{\phi^{-1}y+\gamma\psi}{\phi^{-1}+\gamma}\right)\right] \tag{8}$$

Should the data be hierarchical, with Y_{ij} denoting the jth outcome measured for cluster (subject) i, $i=1,\ldots,N, j=1,\ldots,n_i$ and $\boldsymbol{Y_i}$ the n_i-dimensional vector of all measurements available for cluster i, then the scalar θ_i becomes a vector $\boldsymbol{\theta}_i=(\theta_{i1},\ldots,\theta_{in_i})'$, with $E(\boldsymbol{\theta}_i)=\boldsymbol{\mu}_i$ and $\mathrm{Var}(\boldsymbol{\theta}_i)=\Sigma_i$. In line with the univariate case produces $E(\boldsymbol{Y_i})=\boldsymbol{\mu}_i$ and $\mathrm{Var}(\boldsymbol{Y_i})=M_i+\Sigma_i$, where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the diagonal.

Next, it is possible to include normal random effects in the linear predictor of the generalized linear model, giving rise to the family known as generalized linear mixed model.[7–9,16,17] Assume that, in analogy with (1), conditionally upon q-dimensional random effects $\boldsymbol{b_i} \sim N(\mathbf{0}, D)$, the outcomes Y_{ij} are independent with densities of the form

$$f_i(y_{ij}|\boldsymbol{b_i},\boldsymbol{\xi},\phi)=\exp\left\{\phi^{-1}[y_{ij}\lambda_{ij}-\psi(\lambda_{ij})]+c(y_{ij},\phi)\right\} \tag{9}$$

with

$$\eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) = \eta[E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi})] = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i} \tag{10}$$

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

This kind of models are a bit less common for survival data, where so-called frailty models,[1] rather of the type with conjugate random effects, are more standard. In any case, the next section presents a framework to combine both types of random effects, with focus on time-to-event data.

4 Models combining conjugate and normal random effects

4.1 General model formulation

Combining both the overdispersion and the normal random effects led Molenberghs et al. to the combined model family

$$f_i(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}, \theta_{ij}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\} \tag{11}$$

with notation similar to the one used in (9), but now with conditional mean

$$E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}, \theta_{ij}) = \mu^c_{ij} = \theta_{ij}\kappa_{ij} \tag{12}$$

where the random variable $\theta_{ij} \sim \mathcal{G}_{ij}(\vartheta_{ij}, \sigma^2_{ij}), \kappa_{ij} = g(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}), \vartheta_{ij}$ is the mean of θ_{ij} and σ^2_{ij} is the corresponding variance. Finally, as before, $\boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$. Write $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$. We now need 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 $\boldsymbol{b_i}$, 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 $\mathcal{G}_{ij}(\vartheta_{ij}, \sigma^2_{ij})$ needs to be replaced with a multivariate one; and (3) they are equal to each other, useful in applications with exchangeable outcomes Y_{ij}.

Obviously, parameterization (12) allows for random effects θ_{ij} capturing overdispersion, and formulated directly at mean scale, whereas κ_{ij} could be considered the GLMM component. The relationship between mean and natural parameter now is

$$\lambda_{ij} = h(\mu^c_{ij}) = h(\theta_{ij}\kappa_{ij}) \tag{13}$$

Details and generic expressions are provided in online Appendix A.

Molenberghs et al.[5] set up a framework to describe under what conditions model (11) still allows for conjugacy. They considered conjugacy, conditional upon the normally-distributed random effect $\boldsymbol{b_i}$. To this effect, they wrote (suppressing non-essential arguments from the functions)

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

generalizing (6), and retain (7). Applying the transformation theorem to (7) leads to

$$f(\theta|\gamma,\psi) = \kappa(\kappa\theta|\tilde{\gamma},\tilde{\psi})$$

Next, we request that the parametric form (7) 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 $\tilde{\gamma}$ and $\tilde{\psi}$ upon absorption of κ. Then, the marginal model, in analogy with (8), equals

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

The condition is termed *strong conjugacy*. Fortunately, the Weibull and exponential cases satisfy this property, with gamma random effects. Other examples include the normal and Poisson cases, with normal and gamma random effects, respectively.[5]

4.2 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

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i,\boldsymbol{b}_i) = \prod_{j=1}^{n_i} \lambda\rho\theta_{ij} y_{ij}^{\rho-1} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b}_i} e^{-\lambda y_{ij}^{\rho}\theta_{ij} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b}_i}} \tag{17}$$

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} \theta_{ij}^{\alpha_j-1} e^{-\theta_{ij}/\beta_j} \tag{18}$$

$$f(\boldsymbol{b}_i) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} e^{-\frac{1}{2}\boldsymbol{b}'_i D^{-1}\boldsymbol{b}_i} \tag{19}$$

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, strong conjugacy applies. This is typically considered for the exponential model, but it holds for the Weibull model too, 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 re-writing 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[1] to set $\alpha_j\beta_j = 1$, for reasons of identifiability. In this case, (18) is replaced by

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{\left(\frac{1}{\alpha_j}\right)^{\alpha_j}\Gamma(\alpha_j)} \theta_{ij}^{\alpha_j-1} e^{-\alpha_j\theta_{ij}} \tag{20}$$

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

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \delta_j e^{-\delta_j \theta_{ij}} \tag{21}$$

implying that the gamma density is reduced to an exponential one, of the form (4) with φ now taking the role of $\delta_j = 1/\beta_j$. Closed-form expressions for the marginal density, means, variances, covariances, and moments are derived in online Appendix B, where also a number of related facts are derived.

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

$$f_i(\boldsymbol{y_i}|\boldsymbol{\vartheta}, D, \vartheta_i, \Sigma_i) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \boldsymbol{b_i}, \boldsymbol{\theta}_i) f(\boldsymbol{b_i}|D) f(\boldsymbol{\theta}_i|\vartheta_i, \Sigma_i) \, \mathrm{d}\boldsymbol{b_i} \, \mathrm{d}\boldsymbol{\theta}_i \tag{22}$$

Here, $\boldsymbol{\vartheta}$ groups all parameters in the conditional model for $\boldsymbol{Y_i}$. From (22) the likelihood derives as

$$L(\boldsymbol{\vartheta}, D, \vartheta_i, \Sigma) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \boldsymbol{b_i}, \boldsymbol{\theta}_i) f(\boldsymbol{b_i}|D) f(\boldsymbol{\theta}_i|\vartheta_i, \Sigma_i) \mathrm{d}\boldsymbol{b_i} \mathrm{d}\boldsymbol{\theta}_i \tag{23}$$

The key problem in maximizing (23) is the presence of N integrals over the random effects $\boldsymbol{b_i}$ and $\boldsymbol{\theta}_i$. We consider so-called partial marginalization, in agreement with Molenberghs et al.[5] but, unlike these authors, also allowing for censorship. We further explore the use of pseudo-likelihood as an alternative to full maximum likelihood.

5.1 Partial marginalization

While closed-form expressions, as derived in online Appendix B, can be used to implement maximum likelihood estimation, with numerical accuracy governed by the number of terms included in the Taylor series, one can also proceed by what Molenberghs et al.[5] termed partial marginalization. By this we refer to integrating the conditional density over the gamma random effects only, leaving the normal random effects untouched. The corresponding probability in the Weibull case is

$$f(y_{ij}|\boldsymbol{b_i}) = \frac{\lambda \kappa_{ij} e^{\mu_{ij}} \rho y_{ij}^{\rho-1} \alpha_j \beta_j}{(1 + \lambda \kappa_{ij} e^{\mu_{ij}} \beta_j y_{ij}^{\rho})^{\alpha_j + 1}} \tag{24}$$

Now, in the survival case it is evidently very likely that censoring occurs. Focusing on right-censored data, it is then necessary to integrate the marginal density over the survival time within the interval $[0, C_i]$. The corresponding cumulative distribution is given in (B.9).

In the spirit of (24), the partial marginalization of a censored component takes the form

$$f(C_{ij}|\boldsymbol{b_i}) = \int_{C_{ij}}^{+\infty} f(y_{ij}|\boldsymbol{b_i}) \mathrm{d}y_{ij} = \frac{1}{(1 + \lambda \kappa_{ij} e^{\mu_{ij}} C_{ij}^{\rho})^{\alpha_j}} \tag{25}$$

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.

5.2 Pseudo-likelihood

Pseudo-likelihood,[10,18] as generalized estimating equations[19] are useful when the computational burden of full likelihood becomes burdensome and/or when robustness against misspecification of higher order moments is desirable. This is especially the case when the joint marginal distribution is available but cumbersome to manipulate and evaluate. Essentially then, 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 re-compose 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.

Let us define pseudo-likelihood in general and formally, after which we turn to the special case of pairwise likelihood. Also the term composite likelihood is encountered in this context. Using Arnold and Strauss,[20] we introduce pseudo-likelihood, the principal idea of which is to replace a numerically challenging joint density by a simpler function assembled from suitable factors.

Define S to be the set of all $2^n - 1$ vectors of length n, consisting solely of zeros and ones, with each vector having at least one non-zero entry. Denote by $\boldsymbol{y}_i^{(s)}$ the subvector of $\boldsymbol{y}_i$ corresponding to the components of s that are non-zero. The associated joint density is $f_s(\boldsymbol{y}_i^{(s)}; \boldsymbol{\xi})$. To define a pseudo-likelihood function, one chooses a set $\delta = \{\delta_s | s \in S\}$ of real numbers, with at least one non-zero component. The log of the pseudo-likelihood is then defined as

$$p\ell = \sum_{i=1}^{N} \sum_{s \in S} \delta_s \ln f_s(\boldsymbol{y}_i^{(s)}; \boldsymbol{\xi}) \tag{26}$$

Adequate regularity conditions have to be invoked to ensure that (26) can be maximized by solving the pseudo-likelihood (score) equations, the latter obtained by differentiating the logarithmic pseudo-likelihood and by equating its derivative to zero. More detail can be found in online Appendix C, where the regularity conditions are given. In particular, when the components in (26) result from a combination of marginal and conditional distributions of the original distribution, then a valid pseudo-likelihood function results. More specifically, the classical log-likelihood function is found by setting $\delta_s = 1$ if s is the vector consisting solely of ones, and 0 otherwise. More details can be found in Varin,[21] Lindsay,[22] and Joe and Lee.[23] Note that Joe and Lee[23] use weighting for reasons of efficiency in pairwise likelihood, similar in spirit to Geys et al.,[24] but differently from its use here, which focuses on bias correction when data are incomplete. Another important reference is Cox and Reid.[25]

Let θ_0 be the true parameter. Under suitable regularity conditions (see Refs. 18, 20, and 26), it can be shown[10] that maximizing the function (26) produces a consistent and asymptotically normal estimator $\widetilde{\boldsymbol{\xi}}_0$ so that $\sqrt{N}(\widetilde{\boldsymbol{\xi}}_0 - \boldsymbol{\xi}_0)$ converges in distribution to $N_p[\mathbf{0}, I_0(\boldsymbol{\xi}_0)^{-1} I_1(\boldsymbol{\xi}_0) I_0(\boldsymbol{\xi}_0)^{-1}]$ with $I_0(\boldsymbol{\theta})$ and $I_1(\boldsymbol{\theta})$ defined by their elements

$$I_{0,k_1k_2}(\boldsymbol{\theta}) = -\sum_{s \in S} \delta_s E_{\boldsymbol{\theta}} \left(\frac{\partial^2 \ln f_s(\boldsymbol{y}^{(s)} | \boldsymbol{\theta})}{\partial \theta_{k_1} \partial \theta_{k_2}} \right)$$

$$I_{2,k_1k_2}(\boldsymbol{\theta}) = \sum_{s,t \in S} \delta_s \delta_t E_{\boldsymbol{\theta}} \left(\frac{\partial \ln f_s(\boldsymbol{y}^{(s)} | \boldsymbol{\theta})}{\partial \theta_{k_1}} \frac{\partial \ln f_t(\boldsymbol{y}^{(t)} | \boldsymbol{\theta})}{\partial \theta_{k_2}} \right)$$

As stated earlier, models for non-Gaussian data can become prohibitive when subjected to full maximum likelihood inference, especially with large within-unit replication. le Cessie and van Houwelingen[27] and Geys et al.[24] replace the true contribution of a vector of correlated binary data to the full likelihood, written as $f(y_{i1}, \ldots, y_{in_i})$, by the product of all pairwise contributions $f(y_{ij}, y_{ik})$ $(1 \leq j < k \leq n_i)$, to obtain a pseudo-likelihood function. Also the term *composite likelihood* is encountered in this context. Renard et al.[28] refer to this particular instance of pseudo-likelihood as *pairwise likelihood*. Grouping the outcomes for subject i into a vector $\boldsymbol{Y_i}$, the contribution of the ith cluster to the log pseudo-likelihood then specializes to

$$p\ell_i = \sum_{j<k} \ln f(y_{ij}, y_{ik}) \tag{27}$$

if it contains more than one observation. Otherwise $p\ell_i = f(y_{i1})$. Extension to three-way and higher order pseudo-likelihood is straightforward. All of these are special cases of (26).

6 Marginal distributions and moments

In online Appendix B, along the lines of Molenberghs et al.[5] and Molenberghs and Verbeke,[29] the marginal density and moments are derived. Molenberghs and Verbeke[29] showed that only a finite number of moments is finite. This holds not only for the combined model, but as soon as gamma random effects are combined with Weibull outcomes, i.e., it also applies to the conventional Weibull-gamma model. Because it is possible that even the second and first moments may be infinite, it is wise to check the number of finite moments. Given the moment expression

$$E(Y_{ij}^k) = \frac{\alpha_j B(\alpha_j - k/\rho, k/\rho + 1)}{\lambda^{k/\rho}\beta_j^{k/\rho}} \exp\left(-\frac{k}{\rho}\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \frac{k^2}{2\rho^2}\boldsymbol{z}_{ij}'D\boldsymbol{z}_{ij}\right) \tag{28}$$

with $B(\cdot, \cdot)$ the beta function, it follows that the order $k \leq \alpha_j\rho$ for the corresponding moment to be finite.

7 Analysis of case studies

7.1 Recurrent asthma attacks in children

We will analyze the times-to-event, introduced in Section 2.1. We consider an exponential model, i.e., a model of the form (17) 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 4.

A formal assessment of the treatment effect from all four models is given in Table 5. The treatment effect ξ_1 is stably identifiable in all four models. As can be seen from Table 5, the treatment effects are similar in strength, but including both random effects reduces the evidence, relative to the exponential model. Needless to say that too parsimonious an association structure might lead to liberal test behavior.

Moreover, still considering the combined model, we can also proceed by means of pseudo-likelihood. This is combined with proper inclusion of the censored observations. The model fitting was performed using a SAS macro (available from the authors) in conjunction with the SAS procedure NLMIXED. The result of this analysis can be found in Table 6. Note that the

Table 4. Asthma study: full likelihood.

Effect	Parameter	Exponential Estimate (s.e.)	Exponential-gamma Estimate (s.e.)
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
Effect	Parameter	Exponential-normal Estimate (s.e.)	Combined Estimate (s.e.)
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

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.

Table 5. 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

Table 6. Asthma data: combined model fitted with maximum likelihood and pseudo-likelihood, with and without censoring (model-based s.e.; empirically corrected s.e.).

Effect	Parameter	Full likelihood Estimate (s.e.)	Pseudo-likelihood Estimate (s.e.)
Without censoring			
Intercept	ξ_0	−3.9923 (20.337)	−3.4862 (6.2316; 0.0856)
Treatment effect	ξ_1	−0.0887 (0.0842)	−0.1060 (0.0203; 0.0953)
Shape parameter	λ	0.8130 (16.534)	0.8272 (5.1551; 0.0049)
Gamma parameter	γ	6.8414 (1.7146)	6.7758 (0.6648; 1.1875)
SD random effect	$\sqrt{d}$	0.4720 (0.0416)	0.3958 (0.0202; 0.0383)
With censoring			
Intercept	ξ_0	−4.0195 (28.663)	−3.6233 (0.4998; 0.09381)
Treatment effect	ξ_1	−0.1115 (0.0996)	−0.1269 (0.0221; 0.10571)
Shape parameter	λ	0.7882 (22.592)	0.9189 (0.4590; 0.00003)
Gamma parameter	γ	3.5633 (0.6281)	4.5882 (0.3627; 0.71248)
SD random effect	$\sqrt{d}$	0.5620 (0.0506)	0.4443 (0.0211; 0.03906)

Table 7. Asthma data: combined model fitted with maximum likelihood and pseudo-likelihood, with and without censoring (model-based s.e.; empirically corrected s.e.).

Effect	Parameter	Full likelihood Estimate (s.e.)	Pseudo-likelihood Estimate (s.e.)
Without censoring			
Intercept	ξ_0	−4.1993 (0.0713)	−3.6758 (0.0176; 0.0869)
Treatment effect	ξ_1	−0.0887 (0.0842)	−0.1060 (0.0203; 0.0953)
Gamma parameter	γ	6.8410 (1.7144)	6.7754 (0.6648; 1.1874)
SD random effect	$\sqrt{d}$	0.4721 (0.0416)	0.3958 (0.0202; 0.0383)
With censoring			
Intercept	ξ_0	−4.2575 (0.0833)	−3.7072 (0.0160; 0.0875)
Treatment effect	ξ_1	−0.1116 (0.0996)	−0.1267 (0.0218; 0.1122)
Gamma parameter	γ	3.5634 (0.6282)	4.5833 (0.1747; 0.1895)
SD random effect	$\sqrt{d}$	0.5620 (0.0506)	0.4446 (0.0177; 0.0424)

Table 8. Asthma study: Wald test for treatment effect's assessment in combined model.

Model	*Z*-value	*p*-value
Without censoring full likelihood	−1.0534	0.1461
Without censoring pseudo-likelihood	−1.1123	0.1330
With censoring full likelihood	−1.1205	0.1312
With censoring pseudo-likelihood	−1.1292	0.1294

combined model was conveniently fitted by pseudo-likelihood (specifically pairwise likelihood) based on all the pairs of outcomes within a subject.

It seems that in the result of the combined model (Table 6), there is overdispersion, regardless of whether censoring is taken into account and irrespective of the estimation method, when full likelihood is employed. Note, however, that the standard errors in this case are far from plausible and may point to difficulties with convergence in this case. The pseudo-likelihood methodology does not seem to suffer from this problem. As a result, overdispersion now disappears, given that the standard error values are more trustworthy.

To further address this issue, it might make sense to set the shape parameter equal to one. Re-fitted results for all four models in this way are reported in Table 7. We now find that the standard errors are plausible throughout and that there is no disparity in the overdispersion results. In addition, we performed Wald test for the assessment of treatment effect in the combined model under several different conditions (full likelihood *versus* pseudo-likelihood; with/without censoring), based on the analyses reported in Table 7. The test results are presented in Table 8. The treatment effect ξ_1 is still stably identifiable in all four combined models. It can be seen from Table 8 that the treatment-effect strengths are still similar to the one in combined model treatment effect assessment in Table 5.

Two remarks are in place. First, convergence is reached faster with pseudo-likelihood as opposed to full likelihood. A related finding was reported in Geys et al.[26] where excessive computational requirements could be avoided when using pseudo-likelihood. Second, we noticed that pseudo-likelihood is more robust against the choice of starting values. This is intuitively plausible,

because from a computational standpoint, our pseudo-likelihood behaves as when analyzing bivariate data. The higher the order of the likelihood, the more vulnerable to numerical instabilities.

In conclusion, given that full likelihood nicely converged for the combined models, given that it is best to account for censoring, and given that the shape parameter is redundant, we can consider the bottom left analysis in Table 7 as the final one.

7.2 Survival in cancer patients

Let us fit to the data introduced in Section 2.2, the generalized log-logistic model

$$f(y) = \frac{\varphi\rho y^{\rho-1}\alpha^{\alpha+1}}{(\alpha+\varphi y^{\rho})^{\alpha+1}}$$

generalized logistic model

$$f(y) = \frac{\rho\varphi e^{\rho y}}{(1+\varphi\beta e^{\rho y})^{1/\beta+1}}$$

and generalized Cauchy model

$$f(y) = \frac{1}{\pi}\cdot\frac{\gamma\rho|y|^{\rho-1}}{\gamma^2+|y|^{2\rho}}$$

In the former two cases, we set

$$\begin{aligned}\kappa_i = \exp[&\beta_1 I(T_i=1)+\beta_2 I(T_i=2)\\ &+\beta_3 I(T_i=3)+\beta_4 I(T_i=4)+\beta_5 I(T_i=5)]\end{aligned}$$

where i is the patient index and cancer type $T_i = 1,\ldots,5$ for stomach, bronchus, colon, ovarian, and breast cancer, respectively. For the generalized Cauchy model, predictor function φ is set equal to γ instead. Parameter estimates are presented in Table 9. Model fitting is performed using the SAS procedure NLMIXED. The code can be obtained from the authors, upon request.

Table 9. Parameter estimates (standard errors) for generalized log-logistic, generalized logistic, and generalized Cauchy models, respectively, fitted to the survival data in cancer patients.

Effect	Parameter	Gen. log-logistic	Gen. logistic	Gen. Cauchy
Stomach	β_1	−96.789 (75.740)	−96.792 (75.612)	11.095 (1.353)
Bronchus	β_2	−90.607 (70.403)	−90.610 (70.284)	11.311 (1.339)
Colon	β_3	−90.607 (70.403)	−90.606 (70.282)	12.426 (1.450)
Ovary	β_4	−133.02 (105.80)	−133.02 (105.62)	12.699 (1.553)
Breast	β_5	−95.396 (84.817)	−95.399 (74.690)	13.664 (1.614)
Weibull parameter	ρ	29.220 (23.571)	29.221 (23.532)	
Gamma parameter	α	0.014 (0.0116)	0.014 (0.0116)	
Shape parameter	ρ			7.063 (0.802)
No. of finite moments	k	0	all	7

Table 10. Results of the simulation study, with 10% censored observation.

	Method	Full likelihood				Pseudo-likelihood			
Subjects	Parmeter	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5
50	Estimate	2.5692	−0.0477	1.0003	0.9056	2.0832	−0.0401	1.1245	1.9200
	Mean (s.e.)	0.2937	0.3843	0.2391	0.2351	0.2671	0.3134	0.2400	0.1774
	Bias	0.5692	−0.1477	0.8003	0.4056	0.0832	−0.1401	0.9245	1.4200
	Relative bias	0.2846	−1.4770	4.0015	0.8112	0.0416	−1.4011	4.6226	2.8400
100	Estimate	2.4243	−0.0182	0.8328	0.8955	2.0037	−0.0076	0.9596	1.8588
	Mean (s.e.)	0.2027	0.2653	0.1078	0.1555	0.1936	0.2224	0.1225	0.1256
	Bias	0.4243	−0.1182	0.6328	0.3955	0.0037	−0.1076	0.7596	1.3588
	Relative bias	0.2121	−1.1824	3.1639	0.7909	0.0018	−1.0766	3.7982	2.7175
200	Estimate	2.0476	0.0031	0.8878	0.7459	1.7850	0.0015	1.0493	1.6035
	Mean (s.e.)	0.1281	0.1666	0.0697	0.0984	0.1299	0.1499	0.0593	0.0861
	Bias	0.0476	−0.0968	0.6878	0.2459	−0.2150	−0.0985	0.8493	1.1035
	Relative bias	0.0238	−0.9686	3.4389	0.4918	−0.1075	−0.9851	4.2467	2.2070

A few remarks are worth making. First, the parameters of the log-logistic and generalized logistic are similar, given that the two families are in one-to-one relationship through the logarithmic transformation on the one hand, and the fact that in the second case the data are, of course, log-transformed, on the other. Second, the previous observation notwithstanding, while $\rho\alpha = 0.4178 < 1$, thence no finite moments exist, in the generalized logistic case all moments are finite! Third, in the generalized Cauchy case, there are seven finite moments, implying, of course, that there is a finite mean and a finite variance.

The key scientific question is directed toward the difference in survival across cancer types. The null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ can be tested by means of an approximate $F_{4,64}$ test statistic. For the generalized log-logistic and generalized logistic distributions, we obtain $F = 0.36$ ($p = 0.8344$), while for the generalized Cauchy, we obtain $F = 5.04$ ($p = 0.0013$). The difference is enormous and arguably, can be ascribed to the lack of finite moments in the latter case.

Thus, our analysis illustrates the occurrence, in real life, of distributions without finite moments, with all moments finite, and with a finite number of finite moments. The first one is the more acute one, and it is precisely this one that corresponds to the Weibull-Gamma frailty model, providing an example where the usual regularity conditions are called into question.

In conclusion, we evidently discard the generalized log-logistic analysis for the lack of finite mean and variance. The generalized Cauchy model has a finite mean and finite variance and provides a parsimonious description, unlike the generalized logistic model, in spite of its full series of finite moments. Hence, the generalized Cauchy is our preferred choice to summarize the structure in the data.

8 Simulation study

In this simulation study, we aim to evaluate the performance of the combined model, Weibull model with gamma frailties and normal random effects, under full likelihood and pseudo-likelihood. The design of the simulation study was carried out under different settings, to investigate the impact of sample size, censoring percentage, and method of estimation.

Table 11. Results of the simulation study, with 25% censored observation.

	Method	Full likelihood				Pseudo-likelihood			
Subjects	Parameter	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5
50	Estimate	1.8312	0.0019	0.5892	0.7786	1.4798	−0.0052	0.6810	1.7775
	Mean (s.e.)	0.2734	0.3534	0.0831	0.2332	0.2663	0.3113	0.0991	0.1828
	Bias	−0.1688	−0.0981	0.3892	0.2786	−0.5202	−0.1052	0.4810	1.2775
	Relative bias	−0.0844	−0.9808	1.9461	0.5571	−0.2601	−1.0518	2.4050	2.5550
100	Estimate	1.6196	0.0180	0.6484	0.7068	1.3118	0.0167	0.7434	1.6544
	Mean (s.e.)	0.1809	0.2337	0.0687	0.1693	0.1825	0.2109	0.0811	0.1241
	Bias	−0.3804	−0.0820	0.4484	0.2068	−0.6882	−0.0832	0.5434	1.1544
	Relative bias	−0.1902	−0.8200	2.2419	0.4135	−0.3441	−0.8325	2.7168	2.3088
200	Estimate	1.3714	0.0046	0.7475	0.6265	1.1153	0.0019	0.8631	1.4733
	Mean (s.e.)	0.1176	0.1510	0.0608	0.1062	0.1214	0.1392	0.0718	0.0797
	Bias	−0.6286	−0.0954	0.5475	0.1265	−0.8847	−0.0981	0.6631	0.9733
	Relative bias	−0.3143	−0.9542	2.7374	0.2530	−0.4424	−0.9813	3.3157	1.9467

Table 12. Results of the simulation study, with 50% censored observation.

	Method	Full likelihood				Pseudo-likelihood			
Subjects	Parameter	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5	β_0 2	β_1 0.1	γ 0.2	$\sqrt{d}$ 0.5
50	Estimate	0.8348	0.0108	0.4746	0.6567	0.4517	0.0217	0.5651	1.3166
	Mean (s.e.)	0.2722	0.3467	0.0914	0.2620	0.2734	0.3231	0.1295	0.2030
	Bias	−1.1651	−0.0892	0.2746	0.1567	−1.5483	−0.0783	0.3651	0.8166
	Relative bias	−0.5826	−0.8920	1.3732	0.3135	−0.7742	−0.7832	1.8255	1.6333
100	Estimate	0.7614	−0.0066	0.5663	0.6307	0.4220	−0.0107	0.6472	1.2124
	Mean (s.e.)	0.1856	0.2352	0.1116	0.1825	0.1883	0.2246	0.1379	0.1428
	Bias	−1.2386	−0.1066	0.3663	0.1307	−1.5780	−0.1107	0.4472	0.7124
	Relative bias	−0.6193	−1.0662	1.8314	0.2614	−0.7890	−1.1071	2.2359	1.4248
200	Estimate	0.7312	0.0092	0.5395	0.6356	0.3961	0.0057	0.6167	1.0012
	Mean (s.e.)	0.1300	0.1647	0.0624	0.1283	0.1306	0.1555	0.0742	0.0991
	Bias	−1.2688	−0.0908	0.3395	0.1356	−1.6039	−0.0942	0.4167	0.5012
	Relative bias	−0.6344	−0.9076	1.6977	0.2712	−0.8019	−0.9425	2.0834	1.0025

We used two sets of true parameters, similar in spirit to the ones in Table 7, without and with censoring (full likelihood). The true parameters are not exactly equal to these in Table 7, to avoid convergence issues during the simulation runs. Starting values were chosen so as to reach good convergence properties: $\xi_0 = 2$, $\xi_1 = 0.1$, $\gamma = 0.2$, and $\sqrt{d} = 0.5$. Sample sizes considered were: 50, 100, and 200 subjects. In addition, we generated the number of observations within a subject from a normal $N(\mu = 12, \sigma^2 = 4)$. A censoring covariate is generated from a Bernoulli (π) with $\pi = 0.9$, 0.75, and 0.5, corresponding to 10%, 25%, and 50% of the observations within a

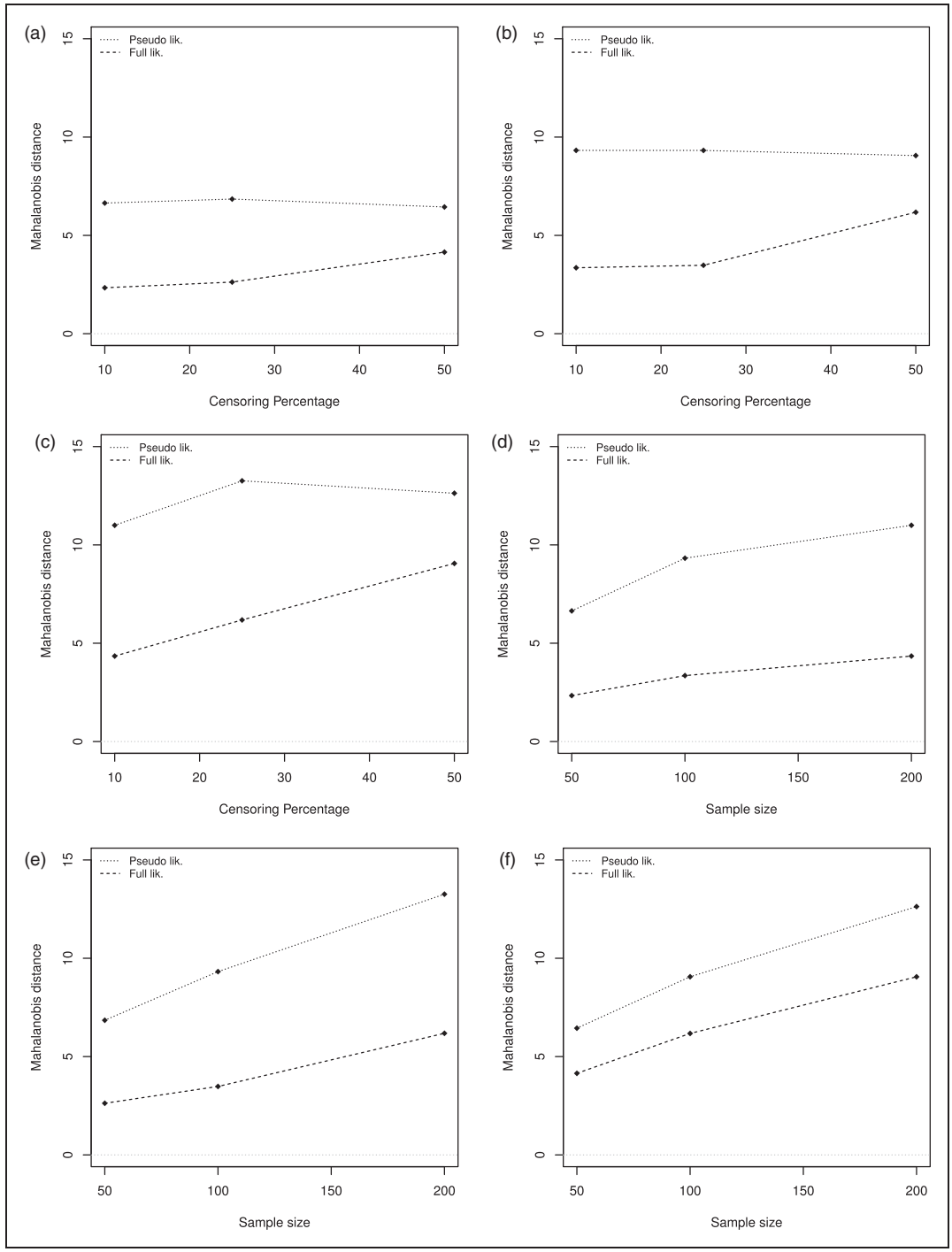

Figure 1. Mahalanobis distance for different sample sizes and censored observation percentages. (a) 50 subjects, (b) 100 subjects, (c) 200 subjects, (d) 10 percent censoring, (e) 25 percent censoring and (f) 50 percent censoring.

subject are censored. This amounts to nine distinct settings, for each of which 500 datasets were generated.

Simulation results are reported in Tables 10, 11, and 12. Each one represents one of the three censoring proportions. The average of the parameter estimates, average of the estimated standard errors of the estimates (mean s.e.), bias, and relative bias are included. The relative bias ranges from −1.47 to 4.62. The proportion of non-converging analysis was relatively small, ranging from 0 to 43 among 500 simulation runs; we found that this proportion increases with censoring.

Furthermore, as a measure of consistency, Mahalanobis distance is used. Precisely, the relative distance between the vector of estimates and the vector of true parameters is computed, for each simulation setting. Some authors use the Euclidian distance as a measure of consistency, including Litière.[30] The Mahalanobis distance has the advance of taking the variance-covariance structure into account. Let $\underset{\sim}{\xi}_0 = (\beta_0, \beta_1, \gamma, \sqrt{d})^T$ represents the vector of the true parameters and $\widehat{\underset{\sim}{\xi}}_0 = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\gamma}, \sqrt{d})^T$ the corresponding vector of estimates, then the Mahalanobis distance is defined by

$$D_M(\widehat{\underset{\sim}{\xi}}_0) = \sqrt{(\widehat{\underset{\sim}{\xi}}_0 - \underset{\sim}{\xi}_0)^T S^{-1} (\widehat{\underset{\sim}{\xi}}_0 - \underset{\sim}{\xi}_0)}$$

where $D_M(\widehat{\underset{\sim}{\xi}}_0)$ denotes Mahalanobis distance and S is the covariance matrix. Based on our simulations, consistency is reached over different sample sizes as well as varying censoring proportions. The set of estimators said to be consistent if the Mahalanobis distance is minimal. Figure 1 displays the evolution of relative distance over increasing proportions of censoring, for a given sample size (panels a, b, and c) and over increasing sample size, for a given censoring proportion (panels d, e, and f). It can be seen in from all panels that pseudo-likelihood estimation method has reduced consistency relative to full likelihood estimation. It can also be observed that, for a given censoring percentage, the relative distances clearly increase with sample size. These occurred for both estimation methods. While with increasing proportion of censored observations, within the same sample size, the relative distance seems to be stable when using pseudo-likelihood estimation method. However, with full likelihood estimation, the relative distance increases as the censoring percentage increases. In other words, there will be loss of some consistency of estimates of the combined model, with increasing censoring percentage, under full likelihood estimation. This result stems from the fact that the pseudo-likelihood method has increased bias in a number of settings, which contributes to the Mahalanobis distance. A similar result was observed by van Duijn et al.[31]

9 Concluding remarks

Building upon work by Molenberghs et al.,[5] we have studied the combination of normal and non-normal random effects in the time-to-event case. We gave particular attention to Weibull models for repeated time-to-event outcomes, with gamma and normal random effects, the so-called combined model. Unlike in the original paper, we allow for right censoring. Furthermore, in line with Molenberghs and Verbeke,[29] we made remarks about the lack of finite moments in the Weibull-gamma model, and hence also in the Weibull-gamma-normal model. On the other hand, the Weibull-gamma-normal model enjoys the so-called *strong conjugacy* property, which is taken to be a version of the well-known conjugacy that is compatible with the additional introduction of normal random effects. This is advantageous when deriving closed-form expressions for the marginal distribution and its corresponding moments.

Whereas Molenberghs et al.[5] confined attention to maximum likelihood estimation, we introduce a pairwise-likelihood version of pseudo-likelihood. Both estimation methods are compared, based on data analysis and simulations. A subtle picture emerges. In a number of cases, maximum likelihood estimation is more efficient in terms of computation time. The statistical loss of efficiency of pseudo-likelihood is relatively small, although the consistency behavior for the maximum-likelihood case is better. That said, pseudo-likelihood has a tremendous advantage in terms of computational stability. Indeed, as illustrated in the data analysis, there are situations where maximum likelihood estimation produces unreliable results due to divergence, no matter what starting values are chosen and other stabilizing measures are taken.

The gamma and normal random effects play distinct roles. In our model formulation, the gamma random effects capture overdispersion, while the normal random effects allow for within-subject association across repeated measures. The model can be extended further and/or adapted to specific cases. For example, when the gamma random effects would be allowed to be correlated from one occasion to the other, then a form of serial (or temporal) association would result. Furthermore, it is possible to generalize the current, two-level formulation, to higher level hierarchies, should this be required.

Computations have been implemented in the SAS procedure NLMIXED, supplemented with user-defined macros. All datasets, software code, and outputs can be found in a WinZip archive on the website www.ibiostat.be/software. Relevant SAS code is also available in online Appendix D.

Funding

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 Conselho de Desenvolvimento Científico e Tecnológico (CNPq), a Brazilian science funding agency.

References

1. Duchateau L and Janssen P. *The frailty model*. New York: Springer, 2007.
2. Hinde J and Demétrio CGB. Overdispersion: models and estimation. *Comput Stat Data Anal* 1998; **27**: 151–170.
3. Nelder JA and Wedderburn RWM. Generalized linear models. *J R Stat Soc Ser B* 1972; **135**: 370–384.
4. McCullagh P and Nelder JA. *Generalized linear models*. London: Chapman & Hall/CRC, 1989.
5. Molenberghs G, Verbeke G, Demétrio C, et al. A family of generalized linear models for repeated measures with normal and conjugate random effects. *Stat Sci* 2010; **25**: 325–347.
6. Molenberghs G, Verbeke G and Demétrio C. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Anal* 2007; **13**: 513–531.
7. Engel B and Keen A. A simple approach for the analysis of generalized linear mixed models. *Stat Neerland* 1994; **48**: 1–22.
8. Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; **88**: 9–25.
9. Wolfinger R and O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *J Stat Comput Simul* 1993; **48**: 233–243.
10. Molenberghs G and Verbeke G. *Models for discrete longitudinal data*. New York: Springer, 2005.
11. Hand DJ, Daly F, Lunn AD, et al. *A handbook of small data sets*, 1st ed. London: Chapman & Hall, 1994, p.255.
12. Breslow N. Extra-Poisson variation in log-linear models. *Appl Stat* 1984; **33**: 38–44.
13. Lawless J. Negative binomial and mixed Poisson regression. *Can J Stat* 1987; **15**: 209–225.
14. Cox DR and Hinkley DV. *Theoretical statistics*. London: Chapman & Hall/CRC, 1974, p.370.
15. Lee Y, Nelder JA and Pawitan Y. *Generalized linear models with random effects: unified analysis via h-likelihood*. Boca Raton, FL: Chapman & Hall/CRC, 2006, p.178.
16. Thall PF and Vail SC. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 1990; **46**: 657–671.
17. Dean CB. Estimating equations for mixed-Poisson models. In: Godambe VP (ed.) *Estimating functions*. Oxford: Oxford University Press, 1991, pp.35–46.
18. Aerts M, Geys H, Molenberghs G, et al. *Topics in modelling of clustered data*. London: Chapman & Hall, 2002.
19. Liang K-Y and Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**: 13–22.
20. Arnold BC and Strauss D. Pseudolikelihood estimation: some examples. *Sankhya B* 1991; **53**: 233–243.
21. Varin C. On composite marginal likelihoods. *Adv Stat Anal* 2008; **92**: 1–28.
22. Lindsay BG. Composite likelihood methods. *Contemp Math* 1988; **80**: 221–239.
23. Joe H and Lee Y. On weighting of bivariate margins in pairwise likelihood. *J Multivariate Anal* 2008; **100**: 670–685.

24. Geys H, Molenberghs G and Lipsitz SR. A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *J Stat Comput Simul* 1998; **62**: 45–72.
25. Cox D and Reid N. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 2004; **91**: 729–737.
26. Geys H, Molenberghs G and Ryan L. Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology. *J Am Stat Assoc* 1999; **94**: 734–745.
27. le Cessie S and van Houwelingen JC. A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics* 1991; **47**: 1267–1282.
28. Renard D, Molenberghs G and Geys H. A pairwise likelihood approach to estimation in multilevel probit models. *Comput Stat Data Anal* 2004; **44**: 649–667.
29. Molenberghs G and Verbeke G. On the Weibull-Gamma frailty model, its infinite moments, and its connection to generalized log-logistic, logistic, Cauchy, and extreme-value distributions. *J Stat Plan Inf* 2011; **141**: 861–868.
30. Litière S. *Random-effects misspecification in generalized linear mixed models: a simulation study*. PhD Thesis report, Chapter 7, University of Hasselt, Diepenbeek, Belgium, 2007.
31. van Duijn MAJ, Gile K and Handcock MS. Comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graphs. *Working paper no. 74, Center for Statistics and the Social Sciences*. Seattle: University of Washington, 2007.