

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

Peer-reviewed author version

MOLENBERGHS, Geert; VERBEKE, Geert; EFENDI, Achmad; BRAEKERS, Roel & Demétrio, Clarice G.B. (2014) A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. In: *Statistical Methods in Medical Research*, 24 (4), p. 434-452.

DOI: 10.1177/0962280214520730

Handle: <http://hdl.handle.net/1942/14833>

# A Combined Gamma Frailty and Normal Random-effects Model for Repeated, Overdispersed Time-to-event Data

Geert Molenberghs<sup>1,2</sup>   Geert Verbeke<sup>2,1</sup>   Achmad Efendi<sup>2,1</sup>  
Roel Braekers<sup>1,2</sup>   Clarice G.B. Demétrio<sup>3</sup>

<sup>1</sup> *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

<sup>2</sup> *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

<sup>3</sup> *ESALQ, Piracicaba, São Paulo, Brazil*

## 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. 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.

**Some 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 (Duchateau and Janssen 2007), it is possible that overdispersion (Hinde and Demétrio 1998) is present in the data, relative to the standard generalized linear model (Nelder and Wedderburn 1972, McCullagh and Nelder 1989) assumed, as well as censored observations.

While each of these features have received attention, it is uncommon to treat all of them simultaneously. Molenberghs *et al.* (2010), building upon Molenberghs, Verbeke, and Demétrio (2007), 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, Engel and Keen 1994, Breslow and Clayton 1993, Wolfinger and O'Connell 1993) for between-subject effects. Here, we supplement their method with the possibility to accommodate censorship.

Whereas Molenberghs *et al.* (2010) focused on maximum likelihood, using so-called partial marginalization, we supplement this inferential option with pairwise likelihood ideas (Molenberghs and Verbeke 2005). 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 (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. Data for the first two patients are listed in Table 1.

### **2.2 Survival in Cancer Patients**

Hand *et al.* (1994, p. 255) 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.

## **3 Background**

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

**Table 1:** *Asthma Data for the First Two Patients. The column labeled 'Status' referred to whether (1) or not (0) censoring has occurred*

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

data. We briefly review these building blocks.

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

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

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

$$E(Y) = \mu = \psi'(\eta), \quad (2)$$

$$\text{Var}(Y) = \sigma^2 = \phi \psi''(\eta), \quad (3)$$

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

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

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

$$\begin{aligned} f(y) &= \varphi \rho y^{\rho-1} e^{-\varphi y^\rho}, \\ E(Y) &= \varphi^{-1/\rho} \Gamma(\rho^{-1} + 1), \\ \text{Var}(Y) &= \varphi^{-2/\rho} \left[ \Gamma(2\rho^{-1} + 1) - \Gamma(\rho^{-1} + 1)^2 \right]. \end{aligned}$$

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  $Y^\rho$ . In the mean and variance expressions for the Weibull,  $\Gamma(\cdot)$  represents the gamma function.

**Table 2:** *Advanced Cancer Data. Survival time in days per patient and per organ affected.*

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

When the standard exponential-family models constrain the mean-variance relationship, so-called overdispersion is introduced. Early reviews are provided by Hinde and Demétrio (1998) provide general treatments of overdispersion. The Poisson case received particular attention by Breslow (1984) and Lawless (1987). A natural step is to allow the overdispersion parameter  $\phi \neq 1$ , so that (3) produces  $\text{Var}(Y) = \phi v(\mu)$ . A convenient route is through a two-stage approach. Generally, the two-stage approach is made up of considering a distribution for the outcome, given a random effect  $f(y_i|\theta_i)$  which, combined with a model for the random effect,  $f(\theta_i)$ , produces the marginal model:

$$f(y_i) = \int f(y_i|\theta_i)f(\theta_i)d\theta_i. \quad (5)$$

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 (Cox and Hinkley 1974, p. 370; Lee, Nelder, and Pawitan 2006, p. 178). 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

**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	$f(\theta)$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$	$\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\beta^\alpha\Gamma(\alpha)}$
Marg. model	$f(y)$	$\frac{\varphi\alpha\beta}{(1+\varphi\beta y)^{\alpha+1}}$	$\frac{\varphi\rho y^{\rho-1}\alpha\beta}{(1+\varphi\beta y^\rho)^{\alpha+1}}$
	$h(\theta)$	$-\theta$	$-\theta$
	$g(\theta)$	$-\ln(\theta)/\varphi$	$-\ln(\theta)/\varphi$
	$\phi$	$1/\varphi$	$1/\varphi$
	$\gamma$	$\varphi(\alpha-1)$	$\varphi(\alpha-1)$
	$\psi$	$[\beta\varphi(\alpha-1)]^{-1}$	$[\beta\varphi(\alpha-1)]^{-1}$
	$c(y, \phi)$	$\ln(\varphi)$	$\ln(\varphi\rho y^{\rho-1})$
	$c^*(\gamma, \psi)$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi) - \ln\Gamma\left(\frac{\gamma+\varphi}{\varphi}\right)$	$\frac{\gamma+\varphi}{\varphi}\ln(\gamma\psi) - \ln\Gamma\left(\frac{\gamma+\varphi}{\varphi}\right)$
Mean	$E(Y)$	$[\varphi(\alpha-1)\beta]^{-1}$	$\frac{\Gamma(\alpha-\rho^{-1})\Gamma(\rho^{-1}+1)}{(\varphi\beta)^{1/\rho}\Gamma(\alpha)}$
Variance	$\text{Var}(Y)$	$\alpha[\varphi^2(\alpha-1)^2(\alpha-2)\beta^2]^{-1}$	$\frac{1}{\rho(\varphi\beta)^{2/\rho}\Gamma(\alpha)} \left[ 2\Gamma(\alpha-2\rho^{-1})\Gamma(2\rho^{-1}) - \frac{\Gamma(\alpha-\rho^{-1})^2\Gamma(\rho^{-1})^2}{\rho\Gamma(\alpha)} \right]$

in the generic forms:

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

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

where  $g(\theta)$  and  $h(\theta)$  are functions,  $\phi$ ,  $\gamma$ , and  $\psi$  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:

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

Should the data be hierarchical, with  $Y_{ij}$  denoting the  $j$ th outcome measured for cluster (subject)  $i$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$  and  $\mathbf{Y}_i$  the  $n_i$ -dimensional vector of all measurements available for cluster  $i$ , then the scalar  $\theta_i$  becomes a vector  $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{in_i})'$ , with  $E(\boldsymbol{\theta}_i) = \boldsymbol{\mu}_i$  and  $\text{Var}(\boldsymbol{\theta}_i) = \boldsymbol{\Sigma}_i$ . In line with the univariate case produces  $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$  and  $\text{Var}(\mathbf{Y}_i) = \mathbf{M}_i + \boldsymbol{\Sigma}_i$ , where  $\mathbf{M}_i$  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 (Thall and Vail 1990, Dean 1991, Engel and Keen 1994, Breslow and Clayton 1993, Wolfinger and O'Connell 1993). Assume that, in analogy with (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

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

with

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

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  $\phi$  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  $\mathbf{b}_i$ .

This kind of models are a bit less common for survival data, where so-called frailty models (Duchateau and Janssen 2007), 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:

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

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

$$E(Y_{ij}|\mathbf{b}_i, \boldsymbol{\xi}, \theta_{ij}) = \mu_{ij}^c = \theta_{ij}\kappa_{ij}, \quad (12)$$

where the random variable  $\theta_{ij} \sim \mathcal{G}_{ij}(\vartheta_{ij}, \sigma_{ij}^2)$ ,  $\kappa_{ij} = g(\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i)$ ,  $\vartheta_{ij}$  is the mean of  $\theta_{ij}$  and  $\sigma_{ij}^2$  is the corresponding variance. Finally, as before,  $\mathbf{b}_i \sim N(\mathbf{0}, D)$ . Write  $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i$ . We now need two different notations,  $\eta_{ij}$  and  $\lambda_{ij}$ , to refer to the linear predictor and/or the natural parameter. The reason is that  $\lambda_{ij}$  encompasses the random variables  $\theta_{ij}$ , whereas  $\eta_{ij}$  refers to the 'GLMM part' only.

It is convenient, but not strictly necessary, to assume that the two sets of random effects,  $\boldsymbol{\theta}_i$  and  $\mathbf{b}_i$ , are independent of each other. Regarding the components  $\theta_{ij}$  of  $\boldsymbol{\theta}_i$ , three useful special cases result from assuming that: (1) they are independent; (2) they are correlated, implying that the collection of univariate distributions  $\mathcal{G}_{ij}(\vartheta_{ij}, \sigma_{ij}^2)$  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 (12) allows for random effects  $\theta_{ij}$  capturing overdispersion, and formulated directly at mean scale, whereas  $\kappa_{ij}$  could be considered the GLMM component. The relationship between mean and natural parameter now is

$$\lambda_{ij} = h(\mu_{ij}^c) = h(\theta_{ij}\kappa_{ij}). \quad (13)$$

Details and generic expressions are provided in Appendix A.

Molenberghs *et al.* (2010) 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  $\mathbf{b}_i$ . To this effect, they wrote (suppressing non-essential arguments from the functions):

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

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

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

Next, we request that the parametric form (7) be maintained:

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

where the parameters  $\gamma^*$  and  $\psi^*$  follow from  $\tilde{\gamma}$  and  $\tilde{\psi}$  upon absorption of  $\kappa$ . Then, the marginal model, in analogy with (8), equals:

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

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 (Molenberghs *et al.* 2010).

## 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

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

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

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

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  $Y_{ij}^\rho$ . 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 (Duchateau and Janssen 2007) to set  $\alpha_j \beta_j = 1$ , for reasons of identifiability. In this case, (18) is replaced by

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

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

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



implying that the gamma density is reduced to an exponential one, of the form (4) with  $\varphi$  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 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

$$f_i(\mathbf{y}_i|\boldsymbol{\vartheta}, D, \boldsymbol{\vartheta}_i, \Sigma_i) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) f(\mathbf{b}_i|D) f(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}_i, \Sigma_i) d\mathbf{b}_i d\boldsymbol{\theta}_i. \quad (22)$$

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

$$L(\boldsymbol{\vartheta}, D, \boldsymbol{\vartheta}_i, \Sigma) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) f(\mathbf{b}_i|D) f(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}_i, \Sigma_i) d\mathbf{b}_i d\boldsymbol{\theta}_i. \quad (23)$$

The key problem in maximizing (23) is the presence of  $N$  integrals over the random effects  $\mathbf{b}_i$  and  $\boldsymbol{\theta}_i$ . We consider so-called partial marginalization, in agreement with Molenberghs *et al.* (2010) 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 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.* (2010) 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:

$$f(y_{ij}|\mathbf{b}_i) = \frac{\lambda\kappa_{ij}e^{\mu_{ij}}\rho y_{ij}^{\rho-1}\alpha_j\beta_j}{(1 + \lambda\kappa_{ij}e^{\mu_{ij}}\beta_j y_{ij}^{\rho})^{\alpha_j+1}}. \quad (24)$$

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:

$$f(C_{ij}|\mathbf{b}_i) = \int_{C_{ij}}^{+\infty} f(y_{ij}|\mathbf{b}_i) dy_{ij} = \frac{1}{(1 + \lambda\kappa_{ij}e^{\mu_{ij}}C_{ij}^{\rho})^{\alpha_j}}. \quad (25)$$

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  $\theta$ , while leaving the normally distributed random effect embedded in the predictor,  $\kappa$  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 (Aerts *et al.* 2002, Molenberghs and Verbeke 2005), as generalized estimating equations (Liang and Zeger 1986) 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 (1991), 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  $\mathbf{y}_i^{(s)}$  the subvector of  $\mathbf{y}_i$  corresponding to the components of  $s$  that are non-zero. The associated joint density is  $f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi})$ . 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

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

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 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 (2008), Lindsay (1988), and Joe and Lee (2008). Note that Joe and Lee (2008) use weighting for reasons of efficiency in pairwise likelihood, similar in spirit to Geys, Molenberghs, and Lipsitz (1998), but differently from its use here, which focuses on bias correction when data are incomplete. Another important reference is Cox and Reid (2004).

Let  $\theta_0$  be the true parameter. Under suitable regularity conditions (see also Arnold and Strauss 1991, Geys, Molenberghs, and Ryan 1999, Aerts *et al.* 2002), it can be shown (Molenberghs and Verbeke 2005) that maximizing the function (26) produces a consistent and asymptotically normal estimator  $\tilde{\boldsymbol{\xi}}_0$  so that  $\sqrt{N}(\tilde{\boldsymbol{\xi}}_0 - \boldsymbol{\xi}_0)$  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:

$$\begin{aligned} I_{0,k_1 k_2}(\boldsymbol{\theta}) &= - \sum_{s \in S} \delta_s E_{\boldsymbol{\theta}} \left( \frac{\partial^2 \ln f_s(\mathbf{y}^{(s)} | \boldsymbol{\theta})}{\partial \theta_{k_1} \partial \theta_{k_2}} \right), \\ I_{2,k_1 k_2}(\boldsymbol{\theta}) &= \sum_{s, t \in S} \delta_s \delta_t E_{\boldsymbol{\theta}} \left( \frac{\partial \ln f_s(\mathbf{y}^{(s)} | \boldsymbol{\theta})}{\partial \theta_{k_1}} \frac{\partial \ln f_t(\mathbf{y}^{(t)} | \boldsymbol{\theta})}{\partial \theta_{k_2}} \right). \end{aligned}$$

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

(1991) and Geys, Molenberghs, and Lipsitz (1998) replace the true contribution of a vector of correlated binary data to the full likelihood, written as  $f(y_{i1}, \dots, y_{in_i})$ , 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, Molenberghs, and Geys (2004) refer to this particular instance of pseudo-likelihood as *pairwise likelihood*. Grouping the outcomes for subject  $i$  into a vector  $\mathbf{Y}_i$ , the contribution of the  $i$ th cluster to the log pseudo-likelihood then specializes to

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

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 Appendix B, along the lines of Molenberghs *et al.* (2010) and Molenberghs and Verbeke (2011), the marginal density and moments are derived. Molenberghs and Verbeke (2011) 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:

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

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  $\xi_1$  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 combined model was conveniently fitted by pseudo-likelihood (specifically pairwise likelihood) based on all the pairs of outcomes within a subject.

**Table 4:** *Asthma Study. Full likelihood. 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.*

Effect	Parameter	Exponential	Exponential-gamma
		Estimate (s.e.)	Estimate (s.e.)
Intercept	$\xi_0$	-3.3709 (0.0772)	-3.9782 (15.354)
Treatment effect	$\xi_1$	-0.0726 (0.0475)	-0.0755 (0.0605)
Shape parameter	$\lambda$	0.8140 (0.0149)	1.0490 (16.106)
Std. dev random effect	$\sqrt{d}$	—	—
Gamma parameter	$\gamma$	—	3.3192 (0.3885)
-2log-likelihood		18,693	18,715

Effect	Parameter	Exponential-normal	Combined
		Estimate (s.e.)	Estimate (s.e.)
Intercept	$\xi_0$	-3.8095 (0.1028)	-3.9923 (20.337)
Treatment effect	$\xi_1$	-0.0825 (0.0731)	-0.0887 (0.0842)
Shape parameter	$\lambda$	0.8882 (0.0180)	0.8130 (16.535)
Std. dev random effect	$\sqrt{d}$	0.4097 (0.0386)	0.4720 (0.0416)
Gamma parameter	$\gamma$	—	6.8414 (1.7146)
-2log-likelihood		18,611	18,629

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 is 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  $\xi_1$  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, Molenberghs, and Ryan (1999) 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,

**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	Pseudo-likelihood
		Estimate (s.e.)	Estimate (s.e.)
<i>Without censoring</i>			
Intercept	$\xi_0$	-3.9923 (20.337)	-3.4862 (6.2316; 0.0856)
Treatment effect	$\xi_1$	-0.0887 (0.0842)	-0.1060 (0.0203; 0.0953)
Shape parameter	$\lambda$	0.8130 (16.534)	0.8272 (5.1551; 0.0049)
Gamma parameter	$\gamma$	6.8414 (1.7146)	6.7758 (0.6648; 1.1875)
Std. dev. random effect	$\sqrt{d}$	0.4720 (0.0416)	0.3958 (0.0202; 0.0383)
<i>With censoring</i>			
Intercept	$\xi_0$	-4.0195 (28.663)	-3.6233 (0.4998; 0.09381)
Treatment effect	$\xi_1$	-0.1115 (0.0996)	-0.1269 (0.0221; 0.10571)
Shape parameter	$\lambda$	0.7882 (22.592)	0.9189 (0.4590; 0.00003)
Gamma parameter	$\gamma$	3.5633 (0.6281)	4.5882 (0.3627; 0.71248)
Std. dev. random effect	$\sqrt{d}$	0.5620 (0.0506)	0.4443 (0.0211; 0.03906)

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:

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

**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	Pseudo-likelihood
		Estimate (s.e.)	Estimate (s.e.)
<i>Without censoring</i>			
Intercept	$\xi_0$	-4.1993 (0.0713)	-3.6758 (0.0176; 0.0869)
Treatment effect	$\xi_1$	-0.0887 (0.0842)	-0.1060 (0.0203; 0.0953)
Gamma parameter	$\gamma$	6.8410 (1.7144)	6.7754 (0.6648; 1.1874)
Std. dev. random effect	$\sqrt{d}$	0.4721 (0.0416)	0.3958 (0.0202; 0.0383)
<i>With censoring</i>			
Intercept	$\xi_0$	-4.2575 (0.0833)	-3.7072 (0.0160; 0.0875)
Treatment effect	$\xi_1$	-0.1116 (0.0996)	-0.1267 (0.0218; 0.1122)
Gamma parameter	$\gamma$	3.5634 (0.6282)	4.5833 (0.1747; 0.1895)
Std. dev. 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

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:

$$\kappa_i = \exp[\beta_1 I(T_i = 1) + \beta_2 I(T_i = 2) + \beta_3 I(T_i = 3) + \beta_4 I(T_i = 4) + \beta_5 I(T_i = 5)],$$

where  $i$  is the patient index and cancer type  $T_i = 1, \dots, 5$  for stomach, bronchus, colon, ovarian, and breast cancer, respectively. For the generalized Cauchy model, predictor function  $\varphi$  is set equal to  $\gamma$  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.

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,

**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	$\beta_1$	-96.789(75.740)	-96.792(75.612)	11.095(1.353)
Bronchus	$\beta_2$	-90.607(70.403)	-90.610(70.284)	11.311(1.339)
Colon	$\beta_3$	-90.607(70.403)	-90.606(70.282)	12.426(1.450)
Ovary	$\beta_4$	-133.02(105.80)	-133.02(105.62)	12.699(1.553)
Breast	$\beta_5$	-95.396(84.817)	-95.399(74.690)	13.664(1.614)
Weibull param.	$\rho$	29.220(23.571)	29.221(23.532)	
Gamma param.	$\alpha$	0.014(0.0116)	0.014(0.0116)	
Shape param.	$\rho$			7.063(0.802)
# finite moments	$k$	0	all	7

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 7 finite moments, implying, of course, that there is a finite mean and a finite variance.

The key scientific question is directed towards 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.

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( $\pi$ ) with  $\pi = 0.9, 0.75$ , and  $0.5$ , corresponding to 10%, 25%, and 50% of the observations within a subject are censored. This amounts to nine distinct settings, for each of which 500 data sets 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 (2007). The Mahalanobis distance has the advance of taking the variance-covariance structure into account. Let  $\xi_0 = (\beta_0, \beta_1, \gamma, \sqrt{d})^T$  represents the vector of the true parameters and  $\widehat{\xi}_0 = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\gamma}, \widehat{\sqrt{d}})^T$  the corresponding vector of estimates, then the Mahalanobis distance is defined by

$$D_M(\widehat{\xi}_0) = \sqrt{(\widehat{\xi}_0 - \xi_0)^T S^{-1} (\widehat{\xi}_0 - \xi_0)},$$

where  $D_M(\widehat{\xi}_0)$  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 word, 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, Gile, and Handcock (2007).

## 9 Concluding Remarks

Building upon work by Molenberghs *et al.* (2010) 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 (2011), 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



**Table 10:** Results of the simulation study, with 10 percent censored observation.

Method		Full.Lik				PL			
Subjects	Par.	$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$	$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$
		2	0.1	0.2	0.5	2	0.1	0.2	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
	Rel.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
	Rel.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
	Rel.Bias	0.0238	-0.9686	3.4389	0.4918	-0.1075	-0.9851	4.2467	2.2070

normal random effects. This is advantageous when deriving closed-form expressions for the marginal distribution and its corresponding moments.

Whereas Molenberghs *et al.* (2010) 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 captures overdispersion, while the normal random effects allows 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](http://www.ibiostat.be/software). Relevant SAS code is also available in Appendix D.

**Table 11:** Results of the simulation study, with 25 percent censored observation.

Subjects	Method	Full.Lik				PL			
		$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$	$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$
	Par.	2	0.1	0.2	0.5	2	0.1	0.2	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
	Rel.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
	Rel.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
	Rel.Bias	-0.3143	-0.9542	2.7374	0.2530	-0.4424	-0.9813	3.3157	1.9467

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

## References

- Aerts, M., Geys, H., Molenberghs, G., and Ryan, L. (2002) *Topics in Modelling of Clustered Data*. London: Chapman & Hall.
- Arnold, B.C. and Strauss, D. (1991) Pseudolikelihood estimation: some examples. *Sankhya B*, **53**, 233–243.
- Breslow, N. (1984) Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38–44.
- Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. London: Chapman & Hall/CRC.
- Cox, D. and Reid, N. (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.
- Dean, C.B. (1991) Estimating equations for mixed-Poisson models. In: *Estimating Functions*, V.P. Godambe (Ed.). Oxford: Oxford University Press.

**Table 12:** Results of the simulation study, with 50 percent censored observation.

Method		Full.Lik				PL			
Subjects	Par.	$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$	$\beta_0$	$\beta_1$	$\gamma$	$\sqrt{d}$
		2	0.1	0.2	0.5	2	0.1	0.2	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
	Rel.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
	Rel.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
	Rel.Bias	-0.6344	-0.9076	1.6977	0.2712	-0.8019	-0.9425	2.0834	1.0025

Duchateau, L. and Janssen, P. (2007) *The Frailty Model*. New York: Springer.

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

Geys, H., Molenberghs, G., and Lipsitz, S.R. (1998) A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *Journal of Statistical Computation and Simulation*, **62**, 45–72.

Geys, H., Molenberghs, G., and Ryan, L. (1999) Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, **94**, 734–745.

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E. (1994). *A Handbook of Small Data Sets* (1st ed.) London: Chapman & Hall.

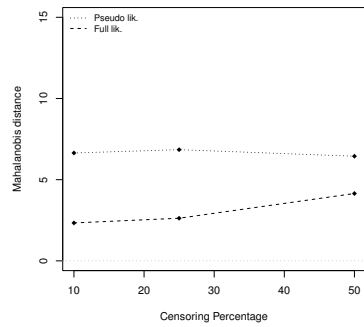
Hinde, J. and Demétrio, C.G.B. (1998) Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.

Joe, H. and Lee, Y. (2008). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, **100**, 670–685.

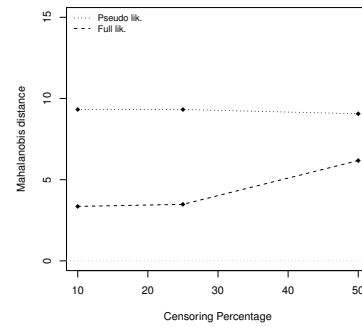
Lawless, J. (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**, 209–225.

le Cessie, S. and van Houwelingen, J.C. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, **47**, 1267–1282.

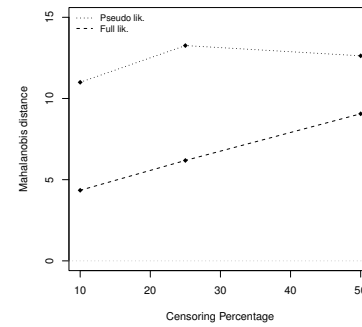
- Lee, Y., Nelder, J.A., and Pawitan, Y. (2006) *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Boca Raton: Chapman & Hall/CRC.
- Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- Lindsay, B.G. (1988) Composite likelihood methods. *Contemporary Mathematics*, **80**, 221–239.
- Litière, S. (2007). Random-effects misspecification in generalized linear mixed models: A simulation study. *PhD thesis report, Chapter 7*. University of Hasselt, Diepenbeek, Belgium.
- McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. London: Chapman & Hall/CRC.
- Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.
- Molenberghs, G. and Verbeke, G. (2011) On the Weibull-Gamma frailty model, its infinite moments, and its connection to generalized log-logistic, logistic, Cauchy, and extreme-value distributions. *Journal of Statistical Planning and Inference*, **141**, 861–868.
- Molenberghs, G., Verbeke, G., and Demétrio, C. (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.
- Molenberghs, G., Verbeke, G., Demétrio, C., and Vieira, A. (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.
- Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series B*, **135**, 370–384.
- Renard, D., Molenberghs, G., and Geys, H. (2004) A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics and Data Analysis*, **44**, 649–667.
- Thall, P.F. and Vail, S.C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.
- van Duijn, M.A.J., Gile, K., and Handcock, M.S. (2007) 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.
- Varin, C. (2008) On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**, 1–28.
- Wolfinger, R. and O’Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233–243.



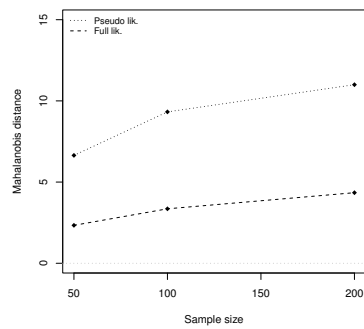
(a) 50 subjects



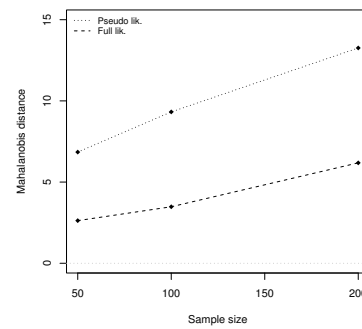
(b) 100 subjects



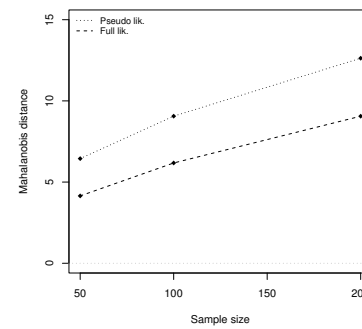
(c) 200 subjects



(d) 10 percent censoring



(e) 25 percent censoring



(f) 50 percent censoring

**Figure 1:** Mahalanobis distance for different sample sizes and censored observation percentages

# A Combined Gamma Frailty and Normal Random-effects Model for Repeated, Overdispersed Time-to-event Data

Geert Molenberghs<sup>1,2</sup> Geert Verbeke<sup>2,1</sup> Achmad Efendi<sup>2,1</sup>  
 Roel Braekers<sup>1</sup> Clarice G.B. Demétrio<sup>3</sup>

<sup>1</sup> I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium

<sup>2</sup> I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

<sup>3</sup> ESALQ, Piracicaba, São Paulo, Brazil

## Supplementary Materials

### A Generic Approximations for Marginal Model Elements

In general, an approximate mean can be derived using the expansion, around  $\mathbf{b}_i = \mathbf{0}$ ,

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

i.e.,

$$E(Y_{ij}) \approx \vartheta_{ij} \left[ g(\eta_{ij}^{(0)}) + \frac{1}{2}g''(\eta_{ij}^{(0)})\mathbf{z}'_{ij}D\mathbf{z}_{ij} \right],$$

where  $\eta_{ij}^{(0)} = \mathbf{x}'_{ij}\boldsymbol{\xi}$ . A general variance expression can be derived in a similar fashion, based upon:

$$\text{Var}(Y_{ij}) = E\{E[\text{Var}(Y_{ij}|\mathbf{b}_i, \theta_{ij})]\} + E\{\text{Var}[E(Y_{ij}|\mathbf{b}_i, \theta_{ij})]\} + \text{Var}\{E[E(Y_{ij}|\mathbf{b}_i, \theta_{ij})]\}.$$

To simplify ensuing derivations, write the variance function as

$$\omega(\mu_{ij}^c) = \omega_{ij}(\theta_{ij}\kappa_{ij}) = \phi\psi''[h(\theta_{ij}\kappa_{ij})]. \quad (\text{A.1})$$

Note that  $\omega(\cdot)$  allows for all of the traditional mean-variance relationships of GLMs for Gaussian, binary, binomial, count, and time-to-event data. Straightforward but tedious algebraic derivations, based on expansions around  $\theta_{ij} = 1$  and  $\mathbf{b}_i = \mathbf{0}$ , leads to:

$$\begin{aligned} \text{Var}(Y_{ij}) \approx & \omega[g(\eta_{ij}^{(0)})] + \omega'[g(\eta_{ij}^{(0)})](\vartheta_{ij} - 1) + \frac{1}{2}\omega''[g(\eta_{ij}^{(0)})]g^2(\eta_{ij}^{(0)})(\sigma_{ij}^2 + \vartheta_{ij}^2 - 2\vartheta_{ij} + 1) \\ & + \frac{1}{2}\omega''[g(\eta_{ij}^{(0)})]g(\eta_{ij}^{(0)})g'(\eta_{ij}^{(0)})(\vartheta_{ij}^2 + \sigma_{ij}^2)\mathbf{z}'_{ij}D\mathbf{z}_{ij} + \sigma_{ij}^2g^2(\eta_{ij}^{(0)}) \\ & + \frac{1}{2}\omega'[g(\eta_{ij}^{(0)})]g''(\eta_{ij}^{(0)})\vartheta_{ij}\mathbf{z}'_{ij}D\mathbf{z}_{ij} + \xi_{ij}^2[g'(\eta_{ij}^{(0)})]^2\mathbf{z}'_{ij}D\mathbf{z}_{ij} \\ & + \sigma_{ij}^2[g'^2(\eta_{ij}^{(0)}) + g(\eta_{ij}^{(0)})g''(\eta_{ij}^{(0)})]\mathbf{z}'_{ij}D\mathbf{z}_{ij}. \end{aligned} \quad (\text{A.2})$$

Likewise, for the covariance function:

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) \approx & \sigma_{ijk} \left[ g(\eta_{ij}^{(0)})g(\eta_{ik}^{(0)}) + \frac{1}{2}g(\eta_{ij}^{(0)})g''(\eta_{ik}^{(0)})\mathbf{z}'_{ik}D\mathbf{z}_{ik} \right. \\ & \left. + \frac{1}{2}g(\eta_{ik}^{(0)})g''(\eta_{ij}^{(0)})\mathbf{z}'_{ij}D\mathbf{z}_{ij} + g'(\eta_{ij}^{(0)})g'(\eta_{ik}^{(0)})\mathbf{z}'_{ij}D\mathbf{z}_{ik} \right] \\ & + \vartheta_{ij}\xi_{ik}g'(\eta_{ij}^{(0)})g'(\eta_{ik}^{(0)})\mathbf{z}'_{ij}D\mathbf{z}_{ik}. \end{aligned} \quad (\text{A.3})$$

Here,  $\sigma_{ijk}$  is the covariance between  $\theta_{ij}$  and  $\theta_{ik}$ . In case these effects are assumed to be independent, a large portion of (A.3) then cancels, with covariance induced solely by the effects  $\mathbf{b}_i$ . In case all  $\theta_{ij}$  are equal,  $\sigma_{ijk} \equiv \sigma_{ij}$ . Evidently, (A.2) and (A.3) lead to approximate expressions for the correlations, too. Of course, in situations where closed forms exist, these expressions need not be used.

Needless to say that the above approximations may or may not be accurate, depending on the context. Therefore, their use should be seen as poor man's choice, when no explicit forms are available. Fortunately, closed forms are available for the normal and Poisson cases (Molenberghs *et al.* 2010), and the Weibull case (Section 4.2) on the other.

## B Marginal Density and Moments

Let us derive the marginal density of the model specified by (17)–(18). First replace the predictor  $\mathbf{x}'_{ij}\boldsymbol{\xi} + \mathbf{z}'_{ij}\mathbf{b}_i$  in (17) by  $\mu$ , and integrate

$$f(y|\theta) = \lambda \rho y^{\rho-1} \theta e^\mu e^{-\lambda y^\rho \theta e^\mu}$$

over the general gamma distribution of  $\theta$ , i.e., over one component of (18):

$$f(y) = \frac{\lambda \rho y^{\rho-1} e^\mu}{\beta^\alpha \Gamma(\alpha)} \int \theta^\alpha e^{-\theta[1/\beta + \lambda y^\rho e^\mu]} d\theta = \frac{\lambda \rho y^{\rho-1} e^\mu \alpha \beta}{(1 + \lambda \beta y^\rho e^\mu)^{\alpha+1}}, \quad (\text{B.1})$$

which easily follows upon setting  $z = 1/\beta + \lambda y^\rho e^\mu$ .

Now, for the general case, first observe that

$$e^{-\lambda y_{ij}^\rho \theta_{ij} e^{\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i}} = \sum_{m_j=0}^{+\infty} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j} y_{ij}^{m_j \rho} \theta_{ij}^{m_j} e^{m_j(\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i)}.$$

It then follows that

$$\begin{aligned} f(\mathbf{y}_i | \boldsymbol{\theta}_i) &= \prod_{j=1}^{n_i} \lambda \rho \theta_{ij} y_{ij}^{\rho-1} \\ &\times \frac{1}{(2\pi)^{q/2} |D|^{1/2}} \int e^{\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i} e^{-\lambda y_{ij}^\rho \theta_{ij} e^{\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i}} e^{-\frac{1}{2} \mathbf{b}_i' D^{-1} \mathbf{b}_i} d\mathbf{b}_i \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} &= \sum_{(m_1, \dots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \theta_{ij}^{m_j+1} \\ &\times \frac{1}{(2\pi)^{q/2} |D|^{1/2}} \int e^{(m_j+1)(\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i) - \frac{1}{2} \mathbf{b}_i' D^{-1} \mathbf{b}_i} d\mathbf{b}_i \end{aligned} \quad (\text{B.3})$$

Now, similar to the binary case (Molenberghs *et al.* 2010), write

$$-\frac{1}{2} \mathbf{b}_i' D^{-1} \mathbf{b}_i + (m_j + 1)(\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i) = -\frac{1}{2} (\mathbf{b}_i - \mathbf{t})' D^{-1} (\mathbf{b}_i - \mathbf{t}) + \ell, \quad (\text{B.4})$$

with

$$\mathbf{t} = (m_j + 1) D \mathbf{z}_{ij}, \quad \ell = (m_j + 1) \left[ \mu_{ij} + \frac{1}{2} (m_j + 1) \mathbf{z}'_{ij} D \mathbf{z}_{ij} \right].$$

Combining (B.3) with (B.4) produces

$$f(\mathbf{y}_i|\boldsymbol{\theta}_i) = \sum_{(m_1, \dots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \theta_{ij}^{m_j+1} e^{(m_j+1)[\mu_{ij} + \frac{1}{2}(m_j+1)\mathbf{z}'_{ij} D \mathbf{z}_{ij}]}. \quad (\text{B.5})$$

Further integration over the gamma distribution produces

$$f(\mathbf{y}_i) = \sum_{(m_1, \dots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \frac{\lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} e^{(m_j+1)[\mu_{ij} + \frac{1}{2}(m_j+1)\mathbf{z}'_{ij} D \mathbf{z}_{ij}]}}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} \cdot I_{j, m_j} \quad (\text{B.6})$$

with

$$I_{j, m_j} = \int \theta_{ij}^{m_j+\alpha_j} e^{-\theta_{ij}/\beta_j} d\theta_{ij} = \beta_j^{m_j+\alpha_j+1} \Gamma(m_j + \alpha_j + 1). \quad (\text{B.7})$$

Plugging (B.6) into (B.7) yields

$$f(\mathbf{y}_i) = \sum_{(m_1, \dots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \frac{\Gamma(\alpha_j + m_j + 1) \beta_j^{m_j+1}}{\Gamma(\alpha_j)} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \times \exp \left\{ (m_j + 1) \left[ \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{1}{2} (m_j + 1) \cdot \mathbf{z}'_{ij} D \mathbf{z}_{ij} \right] \right\}. \quad (\text{B.8})$$

In case censorship applies, it is easy to integrate (B.8) over the interval  $[0, C_{ij}]$  or, in a multivariate fashion, over the cube  $[\mathbf{0}, BC_i]$ :

$$F(\mathbf{C}_i) = \sum_{(m_1, \dots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{(m_j + 1)!} \frac{\Gamma(\alpha_j + m_j + 1) \beta_j^{m_j+1}}{\Gamma(\alpha_j)} \lambda^{m_j+1} C_{ij}^{(m_j+1)\rho} \times \exp \left\{ (m_j + 1) \left[ \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{1}{2} (m_j + 1) \cdot \mathbf{z}'_{ij} D \mathbf{z}_{ij} \right] \right\}. \quad (\text{B.9})$$

Evidently, if censorship applies to some but not all of the times within the vector, then the integration can be restricted to these, and the corresponding contribution will be an amalgamation of components taken from (B.8) and (B.9).

As for the moments, let us first derive the moments based upon (B.1):

$$\begin{aligned} E(Y^k) &= \int_0^{+\infty} \frac{\lambda \rho y^{\rho-1+k} e^{\mu} \alpha \beta}{(1 + \lambda \beta y^{\rho} e^{\mu})^{\alpha+1}} dy \\ &= \frac{\alpha}{\varphi^{k/\rho} \beta^{k/\rho}} \int_1^{+\infty} t^{-\alpha-1} (t-1)^{k/\rho} dt \\ &= \frac{\alpha}{\varphi^{k/\rho} \beta^{k/\rho}} \int_0^1 z^{\alpha-1-k/\rho} (1-z)^{k/\rho} dz \\ &= \frac{\alpha}{\varphi^{k/\rho} \beta^{k/\rho}} B(\alpha - k/\rho, k/\rho + 1), \end{aligned} \quad (\text{B.10})$$

where  $\varphi = \lambda e^{\mu}$ , and the integrator transformations  $t = 1 + \lambda \beta y^{\rho} e^{\mu}$  and  $t = 1/z$  have been used. Now, (B.10) can be used as the  $k$ th moment, conditional upon  $\mathbf{b}_i$ , as follows:

$$E(Y_{ij}^k | \mathbf{b}_i) = \frac{\alpha_j}{\varphi_{ij}^{k/\rho} \beta_j^{k/\rho}} B(\alpha_j - k/\rho, k/\rho + 1), \quad (\text{B.11})$$



where  $\varphi_{ij} = \lambda e^{\mu_{ij} + \mathbf{z}'_{ij} \mathbf{b}_i}$ . The unconditional moment follows as:

$$\begin{aligned} E(Y_{ij}^k) &= \frac{\alpha_j B(\alpha_j - k/\rho, k/\rho + 1)}{\lambda^{k/\rho} e^{\mu_{ij} k/\rho} \beta_j^{k/\rho}} \frac{1}{(2\pi)^{q/2} |D|^{1/2}} \int e^{-\frac{1}{2} \mathbf{b}_i' D^{-1} \mathbf{b}_i - \frac{k}{\rho} \mathbf{z}'_{ij} \mathbf{b}_i} d\mathbf{b}_i \\ &= \frac{\alpha_j B(\alpha_j - k/\rho, k/\rho + 1)}{\lambda^{k/\rho} e^{\mu_{ij} k/\rho} \beta_j^{k/\rho}} e^{\frac{k^2}{2\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}}, \end{aligned} \quad (\text{B.12})$$

where we rewrote

$$-\frac{1}{2} \mathbf{b}_i' D^{-1} \mathbf{b}_i - \frac{k}{\rho} \mathbf{z}_{ij} \mathbf{b}_i = -\frac{1}{2} (\mathbf{b}_i - \mathbf{t})' D^{-1} (\mathbf{b}_i - \mathbf{t}) + \ell,$$

with

$$\mathbf{t} = \frac{k}{\rho} D \mathbf{z}_{ij}, \quad \ell = \frac{1}{2} \frac{k^2}{\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}.$$

From (B.12), we immediately derive, the following moment expression, with mean, variance, and covariance expressions:

$$E(Y_{ij}^k) = \frac{\alpha_j B(\alpha_j - k/\rho, k/\rho + 1)}{\lambda^{k/\rho} \beta_j^{k/\rho}} \exp\left(-\frac{k}{\rho} \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{k^2}{2\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right), \quad (\text{B.13})$$

$$E(Y_{ij}) = \frac{\alpha_j B(\alpha_j - 1/\rho, 1/\rho + 1)}{\lambda^{1/\rho} \beta_j^{1/\rho}} \exp\left(-\frac{1}{\rho} \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{1}{2\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right), \quad (\text{B.14})$$

$$\begin{aligned} \text{Var}(Y_{ij}) &= \frac{\alpha_j}{\lambda^{2/\rho} \beta_j^{2/\rho}} \exp\left(-\frac{2}{\rho} \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{1}{\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right) \\ &\quad \times \left[ B(\alpha_j - 2/\rho, 2/\rho + 1) \exp\left(\frac{1}{\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right) \right. \\ &\quad \left. - \alpha_j B\left(\alpha_j - \frac{1}{\rho}, \frac{1}{\rho} + 1\right)^2 \right], \end{aligned} \quad (\text{B.15})$$

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) &= \frac{\alpha_j \alpha_k}{\lambda^{2/\rho} \beta_j^{1/\rho} \beta_k^{1/\rho}} \exp\left[-\frac{1}{\rho} (\mathbf{x}'_{ij} \boldsymbol{\xi} + \mathbf{x}'_{ik} \boldsymbol{\xi})\right] \\ &\quad \times B\left(\alpha_j - \frac{1}{\rho}, \frac{1}{\rho} + 1\right) B\left(\alpha_k - \frac{1}{\rho}, \frac{1}{\rho} + 1\right) \\ &\quad \times \exp\left[\frac{1}{2\rho^2} (\mathbf{z}'_{ij} D \mathbf{z}_{ij} + \mathbf{z}'_{ik} D \mathbf{z}_{ik})\right] \left[\exp\left(\frac{1}{\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ik}\right) - 1\right]. \end{aligned} \quad (\text{B.16})$$

It is customary, in the standard frailty model (Duchateau and Janssen 2007), to set  $\beta_j = 1/\alpha_j$ , as in (20) for identifiability purposes. The change to (B.8) on the one hand, and to (B.13)–(B.16) is then both evident and minor. Likewise, the exponential version follows from setting  $\rho = 1$ .

It is of interest to explore in a bit more detail the special case induced by (21). Setting  $\alpha_j = 1$  and  $\beta_j = 1/\delta_j$  implies, for (B.13), upon some rewriting:

$$E(Y_{ij}^k) = \frac{\delta_j^{k/\rho} k}{\lambda^{k/\rho}} \Gamma\left(1 - \frac{k}{\rho}\right) \Gamma\left(\frac{k}{\rho}\right) \exp\left(-\frac{k}{\rho} \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{k^2}{2\rho^2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right). \quad (\text{B.17})$$

Reducing the Weibull distribution to the exponential one, i.e., setting  $\rho = 1$ , we find:

$$E(Y_{ij}^k) = \frac{\delta_j^k k}{\lambda^k} \Gamma(1-k) \Gamma(k) \exp\left(-k \mathbf{x}'_{ij} \boldsymbol{\xi} + \frac{k^2}{2} \mathbf{z}'_{ij} D \mathbf{z}_{ij}\right). \quad (\text{B.18})$$

The cases corresponding to (B.17) and, especially, (B.18) deserve further attention. Generally,  $\Gamma(\alpha - k/\rho)$  poses a problem when  $\alpha - k/\rho$  is a negative integer. For simplicity focusing on a single outcome  $Y$  for the case where  $\alpha = 1$  and  $\beta = 1/\delta$  assuming there are no normal random effects, assembling the linear predictor in  $\mu$ , and writing  $\varphi = \lambda e^\mu$ , we find:

$$f(y) = \frac{\varphi \rho y^{\rho-1} \delta}{(\delta + \varphi y^\rho)^2}, \quad (\text{B.19})$$

$$E(Y^k) = \frac{k}{\rho} \left(\frac{\delta}{\varphi}\right)^{k/\rho} \cdot \Gamma(1 - k/\rho) \cdot \Gamma(k/\rho) \quad (\text{B.20})$$

Note that (B.19) provides a family of distributions, special cases of the Weibull-gamma model in Table 3 that we could describe as Weibull-exponential. Considering further the exponential case with  $\rho = 1$ , yields exponential-exponential distributions, with:

$$f(y) = \frac{\varphi \delta}{(\delta + \varphi y)^2}, \quad (\text{B.21})$$

$$E(Y^k) = k \left(\frac{\delta}{\varphi}\right)^k \cdot \Gamma(1-k) \cdot \Gamma(k). \quad (\text{B.22})$$

Clearly, (B.21) defines a family of distributions without finite moments, exactly like but different from the Cauchy distribution, because  $\Gamma(1-k)$  is undefined for  $k = 1, 2, \dots$ . When  $\rho \neq 1$  but fractional, some but not all moments in (B.20) exist, whereas for irrational values of  $\rho$ , all moments in (B.20) are properly defined. Finally, observe that in the general case, there are combinations possible for  $(\alpha, \rho, k)$  that would lead to negative integers and hence undefined moments (B.17).

Specification (B.19) opens the door for the construction of distributions with arbitrarily wide tails, yet finite moments. it suffices to choose  $\rho = 1 + \epsilon$ , with  $\epsilon$  an arbitrarily small irrational value.

Still concentrating on the single-outcome case for notational simplicity, note that the mean-variance relationship, for the Weibull model with normal and gamma random effect, takes the form:

$$\text{Var}(Y) = E(Y)^2 \left[ \frac{2B\left(\alpha - \frac{2}{\rho}, \frac{2}{\rho}\right)}{B\left(\alpha - \frac{1}{\rho}, \frac{1}{\rho}\right)^2} \cdot \rho \cdot e^{d/\rho^2} - 1 \right].$$

When  $\rho = 1$ , this becomes

$$\text{Var}(Y) = E(Y)^2 \left[ \frac{2(\alpha - 1)}{\alpha - 2} e^d - 1 \right].$$

Finally, when also  $d = 0$ , we find:

$$\text{Var}(Y) = E(Y)^2 \left( \frac{\alpha}{\alpha - 2} \right).$$

## C Regularity Conditions for Pseudo-likelihood

Here, we list the required regularity conditions on the density functions  $f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})$ .

**A0** The densities  $f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})$  are distinct for different values of the parameter  $\boldsymbol{\theta}$ .

**A1** The densities  $f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})$  have common support, which does not depend on  $\boldsymbol{\theta}$ .

**A2** The parameter space  $\Omega$  contains an open region  $\omega$  of which the true parameter value  $\boldsymbol{\theta}_0$  is an interior point.

**A3**  $\omega$  is such that for all  $s$ , and almost all  $\mathbf{y}^{(s)}$  in the support of  $\mathbf{Y}^{(s)}$ , the densities admit all third derivatives

$$\frac{\partial^3 f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})}{\partial\theta_{k_1}\partial\theta_{k_2}\partial\theta_{k_3}}.$$

**A4** The first and second logarithmic derivatives of  $f_s$  satisfy

$$E_{\boldsymbol{\theta}} \left( \frac{\partial \ln f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})}{\partial\theta_k} \right) = 0, \quad k = 1, \dots, p,$$

and

$$0 < E_{\boldsymbol{\theta}} \left( \frac{-\partial^2 \ln f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})}{\partial\theta_{k_1}\partial\theta_{k_2}} \right) < \infty, \quad k_1, k_2 = 1, \dots, p.$$

**A5** The matrix  $I_0$ , to be defined in (??), is positive definite.

**A6** There exist functions  $M_{k_1 k_2 k_3}$  such that

$$\sum_{s \in S} \delta_s E_{\boldsymbol{\theta}} \left| \frac{\partial^3 \ln f_s(\mathbf{y}^{(s)}|\boldsymbol{\theta})}{\partial\theta_{k_1}\partial\theta_{k_2}\partial\theta_{k_3}} \right| < M_{k_1 k_2 k_3}(\mathbf{y})$$

for all  $\mathbf{y}$  in the support of  $f$  and for all  $\boldsymbol{\theta} \in \omega$  and  $m_{k_1 k_2 k_3} = E_{\boldsymbol{\theta}_0}[M_{k_1 k_2 k_3}(Y)] < \infty$ .

## D SAS Code for Asthma Data

### D.1 Full Likelihood

```

/*****
OBJECTIVE: to analyze the Asthma dataset using the extended Weibull model with Gamma
           frailty and random normal effects to deal with the longitudinal survival
times.
DATASET: Example 9 of Duchateau & Janssen (2007);
VARIABLE DESCRIPTION:
Patid: Patient ID;
Begin and End: time interval between events for each
patient;
Status: Right censoring indicator (1=Asthma Attack, 0=censored);
Drug: Treatment indicator (1=Drug, 0=Placebo).

```

Author: Achmad Efendi.

\*\*\*\*\*/

```
libname d <mydirectory>;
options nocenter papersize=A4;
%let linkdata=A:\Phd pre-doc\Project PL [NIH]\PL data;
```

```
PROC IMPORT OUT= WORK.asma
DATAFILE= "&linkdata\asma.xls"
DBMS=EXCEL2003 REPLACE ;
Range="asmadel$";
GETNAMES=YES; RUN;
```

```
data asthma; set asma;run;
proc print data=asthma;run;*/
```

```
* Sorting dataset by Drug, in descending way;
proc sort data=asthma;
by descending Drug Patid;run;
```

```
/* Fitting a Weibull regression model */
proc lifereg data=asthma order=data ;
class Drug;
model Time = Drug / distribution=weibull;run;
proc lifereg data=asthma order=data; /* with censoring*/
class Drug;
model Time*Status(0) = Drug / distribution=weibull;run;
```

```
proc sort data=asthma; by Patid;run;
```

```
/* Weibull model, Exponential model as setting rho=1, via NLMIXED */
```

```
proc nlmixed data=asthma tech=newrap;
/*PH model: lamda=rho, rho=gamma, mu=eta*/
parms Beta_0=-3 Beta_1=-0.2 gamma=1;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1);
expeta = exp(eta);
ll = log(rho) + log(gamma) + (gamma-1)*log(Time) + eta - (rho)*(Time**gamma)*expeta;
model Time ~ general(ll);
run;
```

```
/* Weibull model with random effects, Exponential-Normal */
```

```
proc nlmixed data=asthma qpoints=50;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
```

```

ll = log(rho) + log(gamma) + (gamma-1)*log(Time) + eta - rho*(Time**gamma)*expeta;
model Time ~ general(ll);
random b1 ~ normal(0,sigma**2) subject=Patid;
run;

```

```

/* Weibull model with gamma frailty, Exponential-Gamma */

```

```

proc nlmixed data=asthma tech=quanew;
parms Beta_0=-3 Beta_1=-0.2 lambda=1 alpha=1;
      /* Beta starting values from Weibull Model via LIFEREG */
rho=1;
eta = Beta_0 + Beta_1*(Drug=1);
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
run;

```

```

/* Weibull model with gamma frailty and random Effects, Weibull-Gamma-Normal*/

```

```

proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
bounds lambda > 0, alpha > 0;
parms Beta_0=-3 Beta_1=-0.2 lambda = 1 alpha=3.3 sigma=1;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Patid;
run;

```

```

proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000; /*setting lambda=1*/

```

```

bounds alpha > 0;
parms Beta_0=-3 Beta_1=-0.2 alpha=3.3 sigma=1;
rho=1; lambda = 1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Patid;
run;

```

```

/* incorporating censoring*/

```

```

proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
bounds lambda > 0, alpha > 0;
parms Beta_0=-3 Beta_1=-0.11 lambda = 1 alpha=3.3 sigma=1;

```

```

rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
ll = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Patid;
run;

```

```

/*setting lambda=1, and with censoring*/
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
bounds alpha > 0;
parms Beta_0=-3 Beta_1=-0.11 alpha=3.3 sigma=1;
rho=1; lambda = 1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
ll = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Patid;
run;

```

## D.2 Pseudo-likelihood, No censoring

```

/*****
The objective is fitting Weibull-Gamma-Normal model.
Author: Achmad Efendi.
*****/

```

```

libname d <mydirectory>;
options nocenter papersize=A4;
%let linkdata=A:\Phd pre-doc\Project PL [NIH]\PL data;

```

```

PROC IMPORT OUT= WORK.asma
DATAFILE= "&linkdata\asma.xls"
DBMS=EXCEL2003 REPLACE ;
Range="asmadel$";
GETNAMES=YES; RUN;

```

```

data asma1; set asma;
keep patid drug time; run;
proc sort data=asma1; by patid; run;

```

```

%macro pseudosurv(data= ,subject= ,fixed= ,response= );
proc freq data=&data; tables &subject / out=freq noprint; run;
data freq;set freq; keep &subject count;run;
/*getting pairs*/
proc iml;
use freq; read all into f;
      /*subj dose resp*/
use &data; read all into y[colname=coln];
nf=nrow(f);
ny=nrow(y);
do i=1 to nf;
codesubject = f[i,1];
free newdata;
/*getting matrix when subject=i */
do b=1 to ny;
if y[b,1]=codesubject then do;
m=y[b,];
newdata=newdata//m;
end;
end;
ni=nrow(newdata);
npair=ni*(ni-1)/2;
      pairs=J(npair,4,0);
      h=1;
      do j=1 to (ni-1);
      do k=j+1 to ni;
                                /*subject*/
      pairs[h,1]=codesubject;
pairs[h,2]=newdata[1,2];
                                /*one subject is treated with only one dose*/
      pairs[h,3]=newdata[j,3];
pairs[h,4]=newdata[k,3];
      h=h+1;
      end;
      end;
result=result//pairs;
end;
      /*to create longitudinal data, but first deviding the data into two sets*/
x=result;
x1=x[,1]||x[,2]; *subject and treat;
x2=x[,3]||x[,4]; *responses;
npair=nrow(x1); *new npair, after creating pairs;
z1=x1//x1;
                                /*sorting matrix by the first of two columns*/
call sort(z1,{1 2});
z2=shape(x2,npair*2,1);

```

```

/*combining two sets as well as creating the pair numbers*/
a1=1:npair;
a2=a1' || a1';
z3=shape(a2,npair*2,1);
z=z1 || z2 || z3;
create last from z; append from z;
quit;
/*creating pair numbers within subject*/
proc freq data=last;
tables coll / out=freq noprint; run;
data freq;
set freq; keep coll count; run;
proc iml;
use freq; read all into f;
use last; read all into w;
nf=nrow(f);
do i=1 to nf;
a1=1:f[i,2]/2;
a2=a1//a1;
a3=shape(a2',f[i,2],1);
b=b//a3;
end;
c=w || b;
cname = {"Subj", "Fixed", "Resp", "Pairnum", "Pairsubj"};
create last1 from c[colname=cname]; append from c;
quit;
%mend pseudosurv;
%pseudosurv(data=asma1, subject= patid,fixed= drug,response= time);

data asma2; set last1;
patid=subj;
drug=fixed;
time=resp;
keep patid drug time pairnum pairsubj;run;

proc lifereg data=asma2 order=data ;
class drug;
model time = drug / distribution=weibull;run;

/* Weibull model via NLMIXED */

/*PH model: lamda=rho, rho=gamma, mu=eta*/
proc nlmixed data=asma2 tech=newrap;
parms Beta_0=-3.9 Beta_1=-0.09 gamma=1;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1);
expeta = exp(eta);

```



```

ll = log(rho) + log(gamma) + (gamma-1)*log(Time) + eta - (rho)*(Time**gamma)*expeta;
model Time ~ general(ll);
run;

/*****
  Weibull model with gamma frailty and random Effects,
  fitting model that has five parameters in it
*****/

proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
bounds lambda > 0, alpha > 0;
parms Beta_0=-3.3 Beta_1=-0.08 lambda=1 alpha=3.3 sigma=1;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairnum;
ods output ParameterEstimates=initial;
run;

data initial;
set initial;
keep estimate;
run;

proc iml;
use initial; read all into r;
b0=r[1,1];
b1=r[2,1];
b2=r[3,1];
b3=r[4,1];
b4=r[5,1];
call symput('es0',left(char(b0)));
call symput('es1',left(char(b1)));
call symput('es2',left(char(b2)));
call symput('es3',left(char(b3)));
call symput('es4',left(char(b4)));
quit;

data asma3;
set asma2;
drop pairnum;
run;

proc sort data=asma3;

```

```

by patid;
run;

%macro se5;
%do i=1 %to 230;
data asm&i; set asma3; where patid=&i; run;
proc sort data=asm&i; by pairsubj; run;
proc nlmixed data=asm&i tech=quanew qpoints=50 noad maxit=1000 start hess;
bounds lambda > 0, alpha > 0;
parms Beta_0=&es0 Beta_1=&es1 lambda=&es2 alpha=&es3 sigma=&es4;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairsubj;
ods output startingvalues=grad startinghessian=hess&i;
run;
data grad&i; set grad; keep gradient; run;
%end;
data score; set _null_; run;
data hess; set _null_; run;
%do i=1 %to 230;
    data score;set score grad&i;run;
data hess;set hess hess&i;run;
%end;
data hess; set hess; drop row parameter; run;
proc iml;
use score; read all into s;
use hess; read all into h;
nid=nrow(s)/5;
nidh=nrow(h)/5;
a=1:nid; b=J(5,1,1); c=b@(a');
a1=1:nidh; b1=J(5,1,1); c1=b1@(a1');
create s1 from c; append from c;
create s2 from c1; append from c1;
quit;
proc sort data=s1; by col1; run;
proc sort data=s2; by col1; run;
proc iml;
use score; read all into scor;
use hess; read all into hes;
use s1; read all into j1;
use s2; read all into j2;
scor1=j1||scor;
/*combined score functions and hessian matrices*/

```

```

hes1=j2||hes;
/*start summing up all score functions*/
I1=J(5,5,0);
do i=1 to 230;
free g;
do b=1 to (230*5);
if scor1[b,1]=i then do;
m=scor1[b,2];
g=g//m;
/*score functions for one subject*/
end;end;
sg=g*g';
I1=I1+sg;
end;
/*start summing up all hessian matrices*/
I0=J(5,5,0);
do i=1 to 226;
free g1;
do b=1 to (226*5);
if hes1[b,1]=i then do;
m=hes1[b,2]||hes1[b,3]||hes1[b,4]||hes1[b,5]||hes1[b,6];
/*hessian matrices for one subject*/
g1=g1//m;
end;end;
I0=I0+g1;
end;
vmatn=inv(I0);
naiv_se=sqrt(vecdiag(vmatn));
vmat=inv(I0)*I1*inv(I0);
emp_se=sqrt(vecdiag(vmat));
/*naive and empr se*/
se=naiv_se||emp_se;
print se;
quit;
%mend;
%se5;

/*****
Weibull model with gamma frailty and random Effects,
fitting model that has four parameters in it, lambda fixed
*****/

proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
bounds alpha > 0;
parms Beta_0=-3.3 Beta_1=-0.08 alpha=3.3 sigma=1; /*works*/
rho=1; lambda=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;

```

```

expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairnum;
ods output ParameterEstimates=initial2;
run;

data initial2;
set initial2;
keep estimate;
run;

proc iml;
use initial2; read all into r;
b0=r[1,1];
b1=r[2,1];
b2=r[3,1];
b3=r[4,1];
call symput('esa0',left(char(b0)));
call symput('esa1',left(char(b1)));
call symput('esa2',left(char(b2)));
call symput('esa3',left(char(b3)));
quit;

proc sort data=asma3;
by patid;
run;

%macro se4;
%do i=1 %to 230;
data asm&i; set asma3; where patid=&i; run;
proc sort data=asm&i; by pairsubj; run;
proc nlmixed data=asm&i tech=quanew qpoints=50 maxit=1000 start hess;
bounds alpha > 0;
parms Beta_0=&esa0 Beta_1=&esa1 alpha=&esa2 sigma=&esa3;
rho=1; lambda=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairsubj;
ods output startingvalues=grad startinghessian=hes&i;
run;
data grad&i; set grad; keep gradient; run;
%end;

```

```

data score;set _null_; run;
data hess;set _null_; run;
%do i=1 %to 230;
    data score;set score grad&i;run;
    data hess;set hess hes&i;run;
%end;
data hess; set hess; drop row parameter; run;
proc iml;
use score; read all into s;
use hess; read all into h;
nid=nrow(s)/4;
nidh=nrow(h)/4;
a=1:nid; b=J(4,1,1); c=b@(a');
a1=1:nidh; b1=J(4,1,1); c1=b1@(a1');
create s1 from c; append from c;
create s2 from c1; append from c1;
quit;
proc sort data=s1; by col1; run;
proc sort data=s2; by col1; run;
proc iml;
use score; read all into scor;
use hess; read all into hes;
use s1; read all into j1;
use s2; read all into j2;
scor1=j1||scor;
/*combined score functions and hessian matrices*/
hes1=j2||hes;
/*start summing up all score functions*/
I1=J(4,4,0);
do i=1 to 230;
free g;
do b=1 to (230*4);
if scor1[b,1]=i then do;
m=scor1[b,2];
/*score functions for one subject*/
g=g//m;
end;end;
sg=g*g';
I1=I1+sg;
end;
/*summing up all hessian matrices*/
I0=J(4,4,0);
do i=1 to 226;
free g1;
do b=1 to (226*4);
if hes1[b,1]=i then do;
m=hes1[b,2]||hes1[b,3]||hes1[b,4]||hes1[b,5];

```

```

/*hessian matrices for one subject*/
g1=g1//m;
end;end;
I0=I0+g1;
end;
vmatn=inv(I0);
naiv_se=sqrt(vecdiag(vmatn));
vmat=inv(I0)*I1*inv(I0);
emp_se=sqrt(vecdiag(vmat));
/*naive and empr se*/
se=naiv_se||emp_se;
print se;
quit;
%mend;
%se4;

```

### D.3 Pseudo-likelihood, Censoring

```

/*****
The objective is fitting Weibull-Gamma-Normal model.
The data used in this analysis is from recurrent asthma data in children.
Author: Achmad Efendi.
*****/

```

```

libname d <mydirectory>;
options nocenter papersize=A4;
%let linkdata=A:\Phd pre-doc\Project PL [NIH]\PL data;

```

```

PROC IMPORT OUT= WORK.asma
DATAFILE= "&linkdata\asma.xls"
DBMS=EXCEL2003 REPLACE ;
Range="asmadel$";
GETNAMES=YES;RUN;

```

```

data asma1; set asma; keep patid status drug time; run;
proc sort data=asma1; by patid; run;

```

```

/*modified macro for incorporating censorship*/

```

```

%macro plcensor(data= ,subject= ,status= status, fixed= , response= );
proc freq data=&data; tables &subject / out=freq noprint; run;
data freq;set freq; keep &subject count;run;
/*getting pairs*/
proc iml;
use freq; read all into f;

```

```

/*subj dose resp*/
use &data; read all into y[colname=coln];
nf=nrow(f);
ny=nrow(y);
do i=1 to nf;
codesubject = f[i,1];
free newdata;
/*getting matrix when subject=i */
do b=1 to ny;
if y[b,1]=codesubject then do;
m=y[b,];
newdata=newdata//m;
end;end;
ni=nrow(newdata);
Npair=ni*(ni-1)/2;
    pairs=J(Npair,6,0);
    h=1;
    do j=1 to (ni-1);
    do k=j+1 to ni;
/*subject*/
        pairs[h,1]=codesubject;
/*status1, of censoring*/
pairs[h,2]=newdata[j,2];
/*status2, of censoring*/
pairs[h,3]=newdata[k,2];
/*treatment covariate, one subject is treated with only one dose*/
        pairs[h,4]=newdata[1,3];
pairs[h,5]=newdata[j,4];
pairs[h,6]=newdata[k,4];
        h=h+1;
    end;
    end;
result=result//pairs;
end;
/*to create longitudinal data, but first deviding the data into two sets*/
x=result;
x1=x[,1]||x[,4];
x2=x[,2]||x[,3];
x3=x[,5]||x[,6];
z1=x1//x1;
/*sorting matrix by the first of two columns*/
call sort(z1,{1 2});
z2=shape(x2,nrow(x2)*2,1);
z3=shape(x3,nrow(x3)*2,1);
/*combining three sets as well as creating the pair numbers*/
npair=nrow(z1)/2;
a1=1:npair;

```

```

a2=a1//a1;
z4=shape(a2',nrow(z1),1);
z=z1||z2||z3||z4;
create last from z;append from z;
quit;
/*creating pair numbers within subject*/
proc freq data=last;
tables col1 / out=freq noprint; run;
data freq;set freq; keep col1 count; run;
proc iml;
use freq; read all into f;
use last; read all into w[colname=coln];
nf=nrow(f);
do i=1 to nf;
a1=1:f[i,2]/2;
a2=a1//a1;
a3=shape(a2',f[i,2],1);
b=b//a3;
end;
c=w||b;
cname = {"Subj", "Fixed", "Status", "Resp", "Pairnum", "Pairsubj"};
create last1 from c[colname=cname];append from c;
quit;
%mend plcensor;

%plcensor(data=asma1, subject=patid, status=status, fixed=drug, response=time);

data asma2; set last1;
patid=subj;
drug=fixed;
time=resp;
keep patid status drug time pairnum pairsubj;run;

proc sort data=asma2;
by patid;
run;

proc lifereg data=asma2 order=data ;
class drug;
model time = drug / distribution=weibull;
run;

proc lifereg data=asma2 order=data; /* with censoring*/
class Drug;
model Time*Status(0) = Drug / distribution=weibull;
run;

```



```

/*****
Weibull model with gamma frailty and random Effects, incorporating censoring,
fitting model that has five parameters in it.
*****/

proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
bounds lambda > 0, alpha > 0;
parms Beta_0=-3.8 Beta_1=-0.15 lambda=1 alpha=5 sigma=1;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha)**alpha));
c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
ll = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Pairnum;
ods output parameterestimates=initial;
run;

data initial;
set initial;
keep estimate;
run;

proc iml;
use initial; read all into r;
b0=r[1,1];
b1=r[2,1];
b2=r[3,1];
b3=r[4,1];
b4=r[5,1];
call symput('es0',left(char(b0)));
call symput('es1',left(char(b1)));
call symput('es2',left(char(b2)));
call symput('es3',left(char(b3)));
call symput('es4',left(char(b4)));
quit;

data asma3;
set asma2;
drop pairnum;
run;

proc sort data=asma3;
by patid;
run;

```

```

%macro se5;
%do i=1 %to 230;
data asm&i; set asma3; where patid=&i; run;
proc sort data=asm&i; by pairsubj; run;
proc nlmixed data=asm&i tech=quanew qpoints=50 noad maxit=1000 start hess;
bounds lambda > 0, alpha > 0;
parms Beta_0=&es0 Beta_1=&es1 lambda=&es2 alpha=&es3 sigma=&es4;
rho=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
ll = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairsubj;
ods output startingvalues=grad startinghessian=hess&i;
run;
data grad&i; set grad; keep gradient; run;
%end;
data score; set _null_; run;
data hess; set _null_; run;
%do i=1 %to 230;
    data score;set score grad&i;run;
data hess;set hess hess&i;run;
%end;
data hess; set hess; drop row parameter; run;
proc iml;
use score; read all into s;
use hess; read all into h;
nid=nrow(s)/5;
nidh=nrow(h)/5;
a=1:nid; b=J(5,1,1); c=b@(a');
a1=1:nidh; b1=J(5,1,1); c1=b1@(a1');
create s1 from c; append from c;
create s2 from c1; append from c1;
quit;
proc sort data=s1; by col1; run;
proc sort data=s2; by col1; run;
proc iml;
use score; read all into scor;
use hess; read all into hes;
use s1; read all into j1;
use s2; read all into j2;
scor1=j1||scor;
/*combined score functions and hessian matrices*/

```

```

hes1=j2||hes;
/*start summing up all score functions*/
I1=J(5,5,0);
do i=1 to 230;
free g;
do b=1 to (230*5);
if scor1[b,1]=i then do;
m=scor1[b,2];
/*score functions for one subject*/
g=g//m;
end;end;
sg=g*g';
I1=I1+sg;
end;
/*summing up all hessian matrices*/
I0=J(5,5,0);
do i=1 to 226;
free g1;
do b=1 to (226*5);
if hes1[b,1]=i then do;
m=hes1[b,2]||hes1[b,3]||hes1[b,4]||hes1[b,5]||hes1[b,6];
/*hessian matrices for one subject*/
g1=g1//m;
end;end;
I0=I0+g1;
end;
vmatn=inv(I0);
naiv_se=sqrt(vecdiag(vmatn));
vmat=inv(I0)*I1*inv(I0);
emp_se=sqrt(vecdiag(vmat));
se=naiv_se||emp_se; /*naive and empr se*/
print se;
quit;
%mend;
%se5;

/*****
Weibull model with gamma frailty and random Effects, incorporating censoring,
fitting model that has four parameters in it, lambda fixed.
*****/

proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
bounds alpha > 0;
parms Beta_0=-3.8 Beta_1=-0.1 alpha=4.5 sigma=0.2;
rho=1; lambda=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);

```

```

c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha)**alpha);
c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
ll = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=Pairnum;
ods output parameterestimates=initial2;
run;

data initial2;
set initial2;
keep estimate;
run;

proc iml;
use initial2; read all into r;
b0=r[1,1];
b1=r[2,1];
b2=r[3,1];
b3=r[4,1];
call symput('esa0',left(char(b0)));
call symput('esa1',left(char(b1)));
call symput('esa2',left(char(b2)));
call symput('esa3',left(char(b3)));
quit;
proc sort data=asma3; by patid; run;

%macro se4;
%do i=1 %to 230;
data asm&i; set asma3; where patid=&i; run;
proc sort data=asm&i; by pairsbj; run;
proc nlmixed data=asm&i tech=quanew qpoints=50 maxit=1000 start hess;
bounds alpha > 0;
parms Beta_0=&esa0 Beta_1=&esa1 alpha=&esa2 sigma=&esa3;
rho=1; lambda=1;
eta = Beta_0 + Beta_1*(Drug=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=pairsbj;
ods output startingvalues=grad startinghessian=hess&i;
run;
data grad&i; set grad; keep gradient; run;
%end;
data score;set _null_; run;
data hess;set _null_; run;

```

```

%do i=1 %to 230;
    data score;set score grad&i;run;
    data hess;set hess hes&i;run;
%end;
data hess; set hess; drop row parameter; run;
proc iml;
use score; read all into s;
use hess; read all into h;
nid=nrow(s)/4;
nidh=nrow(h)/4;
a=1:nid; b=J(4,1,1); c=b@(a');
a1=1:nidh; b1=J(4,1,1); c1=b1@(a1');
create s1 from c; append from c;
create s2 from c1; append from c1;
quit;
proc sort data=s1; by col1; run;
proc sort data=s2; by col1; run;
proc iml;
use score; read all into scor;
use hess; read all into hes;
use s1; read all into j1;
use s2; read all into j2;
scor1=j1||scor;
/*combined score functions and hessian matrices*/
hes1=j2||hes;
/*summing up all score functions*/
I1=J(4,4,0);
do i=1 to 230;
free g;
do b=1 to (230*4);
if scor1[b,1]=i then do;
m=scor1[b,2];
/*score functions for one subject*/
g=g//m;
end;end;
sg=g*g';
I1=I1+sg;
end;
/*summing up all hessian matrices*/
I0=J(4,4,0);
do i=1 to 226;
free g1;
do b=1 to (226*4);
if hes1[b,1]=i then do;
m=hes1[b,2]||hes1[b,3]||hes1[b,4]||hes1[b,5];
/*hessian matrices for one subject*/
g1=g1//m;

```

```
end;end;
I0=I0+g1;
end;
vmatn=inv(I0);
naiv_se=sqrt(vecdiag(vmatn));
vmat=inv(I0)*I1*inv(I0);
emp_se=sqrt(vecdiag(vmat));
se=naiv_se||emp_se; /*naive and empr se*/
print se;
quit;
%mend;
%se4;
```