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

Peer-reviewed author version

EFENDI, Achmad; MOLENBERGHS, Geert & IDDI, Samuel (2014) A Marginalized Combined Gamma Frailty and Normal Random-effects Model for Repeated, Overdispersed, Time-to-event Outcomes. In: COMMUNICATIONS IN STATISTICS-THEORY AND METHODS, 43 (22), p. 4806-4828.

DOI: 10.1080/03610926.2012.740125
Handle: http://hdl.handle.net/1942/18672

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

Achmad Efendi[1] **Geert Molenberghs**[2,1] **Samuel Iddi**[1]

[1] *I-BioStat, Katholieke Universiteit Leuven, Belgium*

[2] *I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium*

Abstract

This paper proposes a marginalized model for repeated or otherwise hierarchical, overdispersed time-to-event outcomes, adapting the so-called combined model for time-to-event outcomes of Molenberghs *et al* (2012), who combined gamma and normal random effects. The two sets of random effects are used to accommodate simultaneously correlation between repeated measures and overdispersion. The proposed version allows for a direct marginal interpretation of all model parameters. The outcomes are allowed to be censored. Two estimation methods are proposed: full likelihood and pairwise likelihood. The proposed model is applied to data from a so-called comet assay and to data from recurrent asthma attacks in children. Both estimation methods perform very well. From simulation results, it follows that the marginalized combined model behaves similarly to the ordinary combined model in terms of point estimation and precision. It is also observed that the pairwise likelihood required more computation time on the one hand but is less sensitive to starting values and stabler in terms of bias with increasing sample size and censoring percentage than full likelihood, on the other, leaving room for both in practice.

Some Keywords: Combined model; Marginalized multilevel model; Full likelihood; Pairwise likelihood; Weibull distribution.

1 Introduction

Molenberghs *et al* (2010) presented a general framework for modeling (non-)Gaussian overdispersed and hierarchical outcomes. To allow for both overdispersion and correlation simultaneously, they combined conjugate random effects with normal random effects, the latter like in generalized linear mixed models (GLMM; Engel and Keen, 1994; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993). This so-called combined model can flexibly handle all outcomes types commonly encountered in a generalized linear models setting (McCullagh and Nelder, 1989). Here, we focus on time-to-event outcomes, by means of a Weibull distribution with both gamma and normal random effects. A distinctive feature is the potential occurrence of right-censoring, which needs to be incorporated. As such, this model has been studied already in Molenberghs *et al* (2012). These authors fitted their combined Weibull-gamma-normal model with censoring using both full likelihood as well as pairwise likelihood (Molenberghs and Verbeke, 2005). The combined model here is different from the additive frailty model, one of the proposals by Rondeau *et al* (2012), that embeds two correlated random effects, intercept and slope, for proportional hazard models.

The parameters in GLMM are known to have a subject-specific interpretation and not necessarily a population-averaged one. This is perhaps best known for binary data, where often generalized estimating equations (GEE; Liang and Zeger, 1986) are considered as a marginal counterpart to the logistic-normal model. However, GEE lack a likelihood basis, which rules out certain inferential routes and, at the same time, poses challenges when data are incomplete (Molenberghs and Verbeke, 2005). It is therefore useful to make available methodology that simultaneously allows for a marginal as well as a subject-specific interpretation. Depending on the research question, one or the other may be of interest, and sometimes even both. For example, apart from the marginal assessment of treatment effect, one may have an interest in subject-specific effects such as empirical Bayes predictions of random effects, etc. Thus, using both marginal and conditional models at the same time can offer flexibility. Heagerty (1999) and Heagerty and Zeger (2000) proposed a so-called *marginalized multilevel model* (MMM). These authors presented an alternative parameterization for the multilevel model in which the marginal mean, rather than the mean conditional on random effects, is regressed on covariates. An important feature of their approach is that marginal regression parameters are adopted while still permitting individual-level predictions. Furthermore, Griswold and Zeger (2004) reformulated the MMM to render the connection between marginal and conditional models transparent, and then constructed marginalized models in terms of their conditional model counterparts. These authors' formulation did not capture overdispersion (Hinde and Demétrio , 1998), important though it is for non-Gaussian hierarchical data, whereas the combined model does. These considerations lead us to bring together the MMM idea and overdispersion for censored time-to-event outcomes. A related development was made by Iddi and Molenberghs (2012) for binary and count outcomes. As estimation strategies, we use full likelihood estimation with iterative numerical quadrature methods, as well as pairwise likelihood (Molenberghs and Verbeke, 2005). These techniques allow for easy implementation in standard statistical software packages.

The rest of this manuscript is organized as follows. In Section 2, two motivating case studies are described, with their analyses reported in Section 5. The combined model, especially for time-to-event outcomes, is reviewed in Section 3, where we also present our marginalized combined model for time-to-event outcomes. Estimation strategies are the subject of Section 4. A simulation study is reported in Section 6.

2 Motivating Case Study

2.1 Comet Assay Data

A comet assay refers to an easy-to-perform and sensitive technique for the detection of DNA damage at the level of an individual eukaryotic cell. The data used here were collected in four groups of six male rats that received a daily oral dose of a compound in one of three dose levels (low, medium, high) or vehicle control. The treatments were randomized to the rats. On the day of necropsy, an extra group of three animals received a single dose of a positive control (200 mg/kg ethyl methanesulfonate, EMS, PC). The animals were sacrificed three hours after the last dose administration, their liver removed, and processed for the comet assay. A cell suspension was prepared for each animal, from each of which three replicate samples were prepared for scoring. There were 50 randomly selected non-overlapping cells per sample, scored for DNA damage using a semi-automated scoring system. A total of 150 liver cells per animal was scored. DNA damage was assessed through the software system by measuring tail migration, percentage of tail intensity, and tail moment. We are focusing in this paper on the tail-intensity percentage measurement, being the percentage of DNA fragments

Table 1: *Comet data. A part of 50 measurements within the first slide.*

Slide	Tail Intensity	Treatment
1	0.0072	0
1	0.0080	0
1	0.0124	0
1	0.0160	0
1	0.0200	0
1	0.0247	0
1	0.0254	0

present in the tail. The resulting data present themselves in a multilevel structure where a cell suspension or slide, containing three replicate samples, is nested within an animal. In this paper, we use one cluster, slide. Extension is possible but would complicate the illustration aimed at here. In addition, for simpler analysis and focusing on medium-level dose, we target two dose levels (low and medium). As a result, 36 slides, each slide contains roughly 50 measurements, are entered into the analysis. The first seven observation of the first slide can be seen in Table 1.

2.2 Recurrent Asthma Data

The asthma data have been studied in Duchateau and Janssen (2008); they take the form of repeated time-to-event outcomes. Asthma occurs more and more frequently in very young children, i.e., between 6 and 24 months. Therefore, a new application of an existing anti-allergic drug is administered to children who are at higher risk for developing asthma in order to prevent it. A prevention trial is set up with such children randomized to placebo (standard application of the drug) or experimental drug, and the asthma events that developed over time are recorded in a diary. Typically, a patient has more than one asthma event. The intermittent 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 a 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 throughout follow-up, 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 dates of each such risk period are 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 patient are listed in Table 2.

3 Methodology

3.1 The Combined Gamma Frailty and Normal Random Effects Model

For non-Gaussian outcomes, standard exponential-family models are ubiquitous. However, many of these models constrain mean-variance relationship, then raising extra-dispersion concerns. Hinde and Demétrio (1998) provides a general treatment of extra-dispersion. At the same time, hierarchical design, owing to, for example, repeated measures or clustering, have become common. The latter

Table 2: *Asthma data for the first patient. The column labeled 'Status' refers to whether (1) or not (0) censoring occurred.*

Patient ID	Drug	Begin	End	Status	Time
1	0	0	15	1	15
1	0	22	90	1	68
1	0	96	325	1	229
1	0	329	332	1	3
1	0	338	369	1	31
1	0	370	412	1	42
1	0	418	422	1	4
1	0	426	474	1	48
1	0	477	526	1	49
1	0	530	600	0	70

has been addressed, among others, by means of the generalized linear mixed model family (GLMM; Engel and Keen, 1994; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993), which includes normal random effects in the linear predictor. Taking both phenomena simultaneously into account, Molenberghs *et al* (2010) proposed a so-called combined model that brings together normal random effects in the linear predictor with a second set of random effects, usually of a conjugate type, in the sense of Cox and Hinkley (1974, p. 370) and Lee, Nelder, and Pawitan (2006, p. 278).

Assume Y_{ij} is the jth outcome for the ith subject, measured at a time t_{ij}, $(i = 1, \ldots, N; j = 1, \ldots, n_i)$ and assumed to follow an exponential-family distribution, conditional upon random effects $\boldsymbol{b_i}$ and $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{in_i})'$:

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

Here, λ_{ij} is the natural parameter, transforming to the conditional mean $\mu_{ij}^c = E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}_i, \theta_{ij}) = \theta_{ij}\kappa_{ij}$. The components θ_{ij} are assumed independent of one another and follow a distribution

$$\theta_{ij} \sim \mathcal{G}_{ij}(\vartheta_{ij}, \sigma_{ij}^2), \tag{2}$$

with mean ϑ_{ij} and variance σ_{ij}^2. For generality of notation, all aspects of these distribution are subscripted by i and j, whereas in practice one might choose all distributions to be common to a particular measurement occasion j or even common over values of i and of j. In the latter case, for example, the distribution simplifies to $\theta_{ij} \sim \mathcal{G}(\vartheta, \sigma^2)$. Because θ_{ij} enter the mean directly, they need to satisfy the mean's scale. For example, for time-to-event outcomes, θ_{ij} must have support over the positive half line. In contrast, the mean component $\kappa_{ij} = g^{-1}(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i})$ is a conventional GLMM mean component, and hence a function of linear predictor $\eta_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i}$, with link function $g(\cdot)$. For a Weibull distribution, $g^{-1}(\eta) = \exp(\eta)$ is an obvious choice. Further, $\boldsymbol{\xi}$ is the fixed- effects parameter vector and the random effects $\boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$. The advantage of a conjugate choice for θ_{ij} is that not only the range for the mean is respected, but also that closed forms for the marginal mean and variance, and even for the entire marginal distribution, are possible, as shown in Molenberghs *et al* (2010).

Given our focus on time-to-event-outcomes, it is natural to select the Weibull distribution for (1), and then the gamma distribution for (2), because of conjugacy. Assuming its mean $\vartheta_{ij} \equiv \vartheta_j$ and variance $\sigma_{ij}^2 \equiv \sigma_j^2$ constant across subjects and measurements within subjects, and re-parameterizing it using the conventional gamma-distribution parameters, we write $\theta_{ij} \sim \Gamma(\alpha_j, \beta_j)$. This choice is also strongly conjugate in the sense of Molenberghs *et al* (2010), taken to mean that conjugacy still applies even after incorporating normal random effects. With these choices, the model can be represented by its three densities:

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda\rho\theta_{ij} y_{ij}^{\rho-1} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}} e^{-\lambda y_{ij}^{\rho}\theta_{ij} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}}}, \tag{3}$$

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} \theta_{ij}^{\alpha_j-1} e^{-\theta_{ij}/\beta_j}, \tag{4}$$

$$f(\boldsymbol{b_i}) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} e^{-\frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}}, \tag{5}$$

the conditional outcome, conjugate, and normal random effects, respectively. Here, λ and ρ are the conventional additional Weibull scale and shape parameters, respectively. Setting $\rho = 1$ reduces the Weibull to the exponential density. The $\boldsymbol{\xi}$ is fixed effect parameter. In addition, we assume that both conjugate random effect θ_{ij} and normal random effect $\boldsymbol{b_i}$ are independent each other. Model (3)–(5) extends both a GLMM for time-to-event data as well as the gamma frailty model.

3.2 The Marginalized Combined Model

In spite of the combined model's flexibility in accommodating both overdispersion and hierarchical data structures, the fixed-effects vector $\boldsymbol{\xi}$ in (3) is interpreted conditional upon the normal random effects, not marginally, even though this is often desirable. Molenberghs *et al* (2010) show that the model can be marginalized in closed form and in particular that an elegant closed-form exists for the marginal mean function. Consistent with Zeger, Liang, and Albert (1988), they showed that the marginal regression function does not alter, except for the marginal intercept, which depends on the conditional intercept, the scale parameter λ, and the overdispersion parameters α_j and β_j. For inferences regarding the intercepts, or a combination of covariate effects and intercepts, this is cumbersome. While one could proceed by the delta method, its application would be *ad hoc* because specific to every particular model and research question considered. It would require the user to write a piece of code for every particular situation.

To avoid this, we proceed instead by specifying the regression function marginally, with the normal random effects still entering the conditional mean function directly. To this effect, we formulate a *so-called* marginalized multilevel model (MMM), in the tradition of Heagerty (1999), Heagerty and Zeger (2000), and Griswold and Zeger (2004). It takes the form:

$$\mu_{ij}^m = g^{-1}\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi}^m\right), \tag{6}$$

$$\mu_{ij}^c = \theta_{ij}\kappa_{ij} = \theta_{ij} \cdot g^{-1}\left(\Delta_{ij} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i\right). \tag{7}$$

The predictor on the right hand side of (7) replaces the conventional predictor in (3). The fixed-effects parameter vector $\boldsymbol{\xi}^m$ is superscripted to emphasize its directly marginal interpretation. The dual mean specification (6)–(7) immediately leads to a defining equation for connector function Δ_{ij} (Griswold and Zeger, 2004):

$$\begin{aligned} g^{-1}(\boldsymbol{x}_{ij}'\boldsymbol{\xi}^m) = \mu_{ij}^m &= \int_b \int_\theta \theta_{ij} g^{-1}(\Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i) f(\theta_{ij}) f(\boldsymbol{b}_i)\, d\theta_{ij} d\boldsymbol{b}_i \\ &= E(\theta_{ij}) \cdot \int_b g^{-1}(\Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i) f(\boldsymbol{b}_i)\, d\boldsymbol{b}_i. \end{aligned} \tag{8}$$

In the particular case of the Weibull-gamma-normal model, (8) admits a closed-form solution:

$$\Delta_{ij} = -\log(\alpha_j \beta_j) + \boldsymbol{x}_{ij}'\boldsymbol{\xi}^m - \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}/2. \tag{9}$$

This is in contrast to the binary case (Iddi and Molenberghs, 2012). Should there be no gamma random effects, then the first term on the right hand side of (9) simply drops.

4 Estimation

4.1 Full Likelihood Estimation

To conveniently apply full likelihood routines, and given that strong conjugacy applies, we use so-called partial marginalization, proposed in Molenberghs *et al* (2010). The idea is to integrate conditional density (3) analytically over the gamma random effects, leaving the normal random effects to numerical integration. The corresponding marginal-conditional density in the Weibull case is:

$$f(y_{ij}|\boldsymbol{b_i}) = \frac{\lambda e^{\eta_{ij}} \rho y_{ij}^{\rho-1} \alpha_j \beta_j}{(1 + \lambda e^{\eta_{ij}} \beta_j y_{ij}^\rho)^{\alpha_j + 1}}, \tag{10}$$

with $\eta_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i}$ for the conventional formulation and $\eta_{ij} = \Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{b_i}$ for the MMM.

Furthermore, censoring can conveniently be accommodated. With focus on right-censored data, in the spirit of (10), the partial marginalization of a censored component takes the form:

$$f(C_{ij}|\boldsymbol{b_i}) = \int_{C_{ij}}^{+\infty} f(y_{ij}|\boldsymbol{b_i}) dy_{ij} = \frac{1}{(1 + \lambda e^{\eta_{ij}} C_{ij}^\rho)^{\alpha_j}}. \tag{11}$$

Now, having only the normal random effects left, it is then straightforward to use numerical-integration-based full likelihood estimation as implemented in a common statistical software package, such as, for example, the SAS procedure NLMIXED. Indeed, given that the model expressions conditional upon the normal random effect, i.e., (10) and (11), the likelihood function can easily be programmed within the SAS procedure, using the so-called 'general likelihood' feature, which allows for a user-defined model, given normal random effects. The code is provided in Appendix B.

4.2 Pairwise Likelihood Estimation

As an alternative to full likelihood, we also use so-called pairwise likelihood (Renard, Molenberghs, and Geys, 2004). Such an approach can alleviate the cumbersome nature of the joint model expressions needed for full likelihood, especially when there is a lot of within-cluster replication. It can also stabilize computations and make the iterative process less dependent on starting values, even though it may not always reduce computation time. le Cessie and van Houwelingen (1991) and Renard, Molenberghs, and Geys (2004) replace the proper contribution of a vector of correlated binary data to the full likelihood, written as $f(y_{i1},\ldots,y_{in_i})$, by the product of all pairwise contributions $f(y_{ij},y_{ik})$ $(1 \le j < k \le n_i)$, to obtain a so-called pseudo-likelihood function. The contribution of the ith cluster to the log pseudo-likelihood is:

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

This is evidently only for units with more than one observation. Otherwise, $p\ell_i = f(y_{i1})$ is used instead. Pairwise likelihood is a special case of pseudo-likelihood, studied in detail in Arnold and Strauss (1991) and Molenberghs and Verbeke (2005), among others. When it comes to implementation, the ideas of full likelihood are used, but applied to the pairs rather than to the full set of measurements for a particular unit. This leads to a consistent and asymptotically normal parameter estimator, but to obtain valid precision estimates, an information sandwich is required (Molenberghs and Verbeke, 2005).

5 Data Analysis

We consider two examples to which the proposed methodology can usefully be applied. One is censored while the other is not. They exhibit different levels of overdispersion.

5.1 Comet Assay Data

Consider the (uncensored) tail-intensity outcome of the comet data, introduced in Section 2.1. We consider the Weibull-gamma-normal model, i.e., the ordinary combined model, together with its marginalized version. The fixed-effects structure is restricted to a treatment effect. Both full likelihood and pairwise likelihood estimation were used in the model fitting. Parameter estimation was effectuated using the SAS procedure NLMIXED.

Result are presented in Table 3. In line with our earlier comments, it is not surprising that parameter estimates between the marginalized and ordinary combined model are virtually the same, except for the variance of the conjugate random effects, which contributes to the intercept. The treatment effect is highly significant. These considerations hold for both full likelihood and pairwise likelihood.

We see that that the pairwise-likelihood derived standard error estimates are larger than those with full likelihood. This is expected given the well-known optimality of the likelihood estimates. Because of this, and the complete absence of any convergence problems, it is sensible to prefer the full likelihood approach for this data analysis.

5.2 Recurrent Asthma Data

We now turn to the recurrent asthma data, described in Section 2.2. For each of the 226 patients, their treatment allocation and repeated time-to-event outcomes, the time between the end of the previous to onset of the next attack, Y_{ij} is recorded; the outcome is subject to censoring. Also here, the combined model and its marginalized version are considered. Regarding the normal random-effects structure, a random intercept b_{i1} (with variance σ_i^2) and a random slope b_{i2} (with variance σ_e^2) is included. While this could be relaxed, both random effects are assumed to be independently normally distributed. Model fitting is done using both full and pairwise likelihood. Parameter estimates (standard errors) are presented in Table 4.

Table 3: *Comet assay: combined model and the marginalized combined model results. 'WGN' refers to Weibull gamma normal model, the combined model, whilst 'M' means marginalized.*

Effect	Par	WGN	WGN-M	WGN	WGN-M
		Full likelihood		*Pairwise likelihood*	
		Estimate(s.e.)	Estimate(s.e.)	Estimate(s.e.)	Estimate(s.e.)
Treatment	ξ	-2.416(0.086)	-2.464(0.087)	-2.44(0.164)	-2.438(0.162)
Shape	λ	0.234(0.015)	0.252(0.017)	0.239(0.038)	0.239(0.041)
Conj. R.E.	α	52.87(44.16)	20.00(7.115)	24.92(31.71)	27.16(35.84)
s.d. normal R.E.	σ	0.198(0.036)	0.198(0.037)	0.098(0.215)	0.107(0.202)
	-2log-lik	13272	13275	-	-

Full likelihood estimates between the ordinary and marginalized models are similar. Treatment effect is not significant. Because marginalization does not change the likelihood, the likelihood ratios are invariant to this operation (Griswold and Zeger, 2004). Because we now include two normally distributed random effects, the connector function (9) uses a different vector z_{ij}. This now implies that the treatment effect estimate changes upon marginalization, although the change is minor.

Turning attention to results using pairwise likelihood estimation, it is found that the estimates before and after marginalization are still similar. We also see that the estimate of the random slope parameter is virtually zero in all cases, although more pronounced in the pairwise-likelihood case. This does not contradict the results from full likelihood, where this component was non-significant, although the numerical behavior is quite different.

In the four versions presented in the table, the conjugate random effect parameter is statistically significant. This is important and underscores that neither the standard GLMM nor the available marginalized model (Griswold and Zeger, 2004) fits the data adequately.

No convergence problems were encountered. This fact, together with the increased precision for full likelihood in terms of treatment effect and conjugate random effect, it would be sensible here as well to prefer full likelihood. The picture for the normal random effects is slightly different, but the boundary estimate for the standard deviation of the random slope urges caution.

Of course, once should be careful not to over-interpret the results derived from a couple of data analyses. Therefore, we turn to a simulation study.

6 Simulation Study

A simulation study was conducted to evaluate the performance of the Weibull-gamma-normal (WGN) model, as well as its marginalization (WGN-M). We chose the following true parameter values to generate data: $\xi_1 = 2$, $\lambda = 0.2$, $\alpha = 2$, and $\sqrt{d} = 0.25$. ξ_1 is the effect of a covariate, generated from $N(0, 0.1)$. We set $\rho = 1$, turning the Weibull distribution into an exponential one. The design is reminiscent of the asthma data. The true parameter is also used as starting values for model fitting.

Various settings were considered to investigate the impact of different sample sizes (20, 40, 60, and 80

Table 4: *Asthma study. Original and marginalized combined model results. 'WGN' refers to the Weibull-gamma-normal model, whilst 'C' and 'CM' means censored and censored-marginalized, respectively.*

Effect	Par	WGN-C	WGN-CM	WGN-C	WGN-CM
		Full likelihood		*Pairwise likelihood*	
		Estimate(s.e.)	Estimate(s.e.)	Estimate(s.e.)	Estimate(s.e.)
Treatment	ξ	-0.113(0.106)	-0.111(0.102)	-0.127(0.105)	-0.127(0.105)
Shape	λ	0.014(0.001)	0.017(0.001)	0.025(0.002)	0.027(0.003)
Conj.RE	α	3.566(0.632)	3.566(0.632)	4.583(0.708)	4.584(0.708)
s.d. norm. R.int.	σ_i	0.560(0.068)	0.560(0.068)	0.445(0.039)	0.445(0.039)
s.d. norm. R.eff.	σ_e	0.077(0.734)	0.077(0.741)	11E-4(11E-4)	20E-6(20E-6)
-2log-lik		16649	16649		

subjects), with 10 observations per subject, and different right-censoring percentages, i.e., 0, 10, 25, or 50 percent of the simulated time-to-event outcomes. Of course, we could expand the simulation settings to variable numbers of observations per subject, but this would considerably increase the size of the simulation study. We assume independent censoring. This leads to 16 scenarios, for each one of which 500 replications were generated. To each set of data, both the WGN and the WGN-M were fitted, using both full likelihood and pairwise likelihood. In analogy with Molenberghs *et al* (2012), we use the Mahalanobis distance to quantify the bias. Let $\boldsymbol{\xi}_0 = (\xi_1, \lambda, \alpha, \sqrt{d})'$ represent the true parameter vector and $\widehat{\boldsymbol{\xi}_0} = (\widehat{\xi}_1, \widehat{\lambda}, \widehat{\alpha}, \widehat{\sqrt{d}})'$ the estimates, then the Mahalanobis distance D_M is:

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

and S is the covariance matrix. This quantity is straightforward to evaluate because the covariance matrix S follows as a by-product of the optimization process.

The simulation results are reported in Tables A.1–A.8. The average of the parameter estimates, average of standard error estimate (s.e.), the standard error of the estimates (s.e.MC), bias, and relative bias (Rbias) are reported. The bias of the estimates of the WGN are calculated from the true parameters while the bias for the WGN-M are computed from the corresponding marginalization.

The percentage of non-convergence, unsurprisingly, increases with censoring but reduces with higher number of sample sizes. Pairwise likelihood has a slight beneficial impact on convergence, at the cost of increased computation time. Computation time further increases with sample size and with censoring. Whereas the choice between standard or marginalized combined model has no impact on the computation time. The times for fitting a set of data in Tables A.1, A.2, A.3, and A.4, on average are approximately 5 seconds, 30 seconds, 10 seconds, and 1 minute, respectively, using a VSC cluster computer device.

It also follows that the relative bias of the parameter estimates in all simulation settings is rather small, except with the gamma-distributed conjugate random effect, though this disappears with increasing sample size and decreasing censoring percentages. The model based and Monte Carlo standard errors are in good agreement, as one would expect. Relative bias is rather insensitive to estimation method. All of these observations are summarized visually in the Mahalanobis distance (Figure A.1).

Indeed, while the pairwise likelihood produces a higher Mahalanobis distance, increasing the sample sizes gradually removes this effect. We also see that the relative bias of all standard model settings behaves similarly with different percentage of censoring. Finally, with increasing sample size and censorship percentage, the Mahalanobis distance with full likelihood also tends to be higher whereas with pairwise likelihood it appears to be stable.

Additionally, the Mahalanobis distance summarized in Figure A.2 was presented to investigate the bias of the marginalized model from its standard form. It is shown that bias of models using pairwise likelihood is higher than with full likelihood, but then the gap reduced with increasing sample size. We also see that the bias increases with censoring. This behavior occurs in all four different sample size settings.

7 Concluding Remarks

For time-to-event outcomes, we have shown that the combined model (Molenberghs *et al*, 2010, 2012) can be reformulated such that the parameters maintain a marginal interpretation. For this, we proposed a marginalized combined model, for hierarchically organized, overdispersed time-to-event outcomes. The marginalized combined model not only can capture the hierarchical structure in the data and accommodate overdispersion, it also admits a marginal interpretation. Such interpretation is what is needed in many studies set up to answer a (sub)population level question.

Practically, the marginalized multilevel model idea (Heagerty, 1999; Heagerty and Zeger, 2000) is adapted to this context. It is based on specifying marginal and conditional means separately and then linking them by a so-called connector function. Iddi and Molenberghs (2012) followed a similar approach for binary data. Censoring is also allowed for.

The model fitting for the marginalized combined model is straightforward and can be done easily through common statistical software. We made use of the SAS procedure NLMIXED. Both the standard and marginalized versions are easy to fit. Full likelihood and pairwise likelihood were considered, implemented, used for data analysis, and put to the test in a simulation study. The two model versions show similar behavior, across a range of sample sizes, censoring percentages, and estimation methods. In line with findings in Molenberghs *et al* (2012), the pairwise likelihood gives stabler bias than full likelihood, when censoring percentages and sample sizes increase.

Regarding censoring, we restricted attention to right-censored outcomes, but extension of the proposed methodology to left-censoring and interval-censoring is immediate.

Acknowledgement

The authors gratefully acknowledge the financial support from the IAP research network Phase VI of the Belgian Science Policy.

References

Arnold, B.C. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Sankhya B*, **53**, 233–243.

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). *Theoritical Statistics*. London: Chapman & Hall/CRC.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. New York: Springer.

Efendi, A. and Molenberghs, G. (2012). A multilevel model for hierarchical, repeated, and overdispersed time-to-event outcomes and its estimation strategies. *Submitted for publication*

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 Ryan, L.M. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, **94**, 734–745.

Griswold, M.E. and Zeger, S.L. (2004). *On Marginalized multilevel models and their computation*. John Hopkins University, Dept. of Biostatistics Working Papers.

Heagerty, P.J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688–698.

Heagerty, P.J. and Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science*, **15**, 1–26.

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

Iddi, S. and Molenberghs, G.(2012). A combined overdispersed and marginalized multilevel model. *Computational Statistics and Data Analysis*, **56**, 1944–1951.

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*. London: Chapman & Hall/CRC.

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

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., Verbeke, G., Demetrio C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Molenberghs, G., Verbeke, G., Efendi, A., Braekers, R., and Demétrio, C.G.B. (2012). A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *In press.*

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

Rondeau, V., Mazroui, Y., and Gonzalez, J.R. (2012). *Package 'frailtypack'*. CRAN: R project.

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

Zeger, S.L., Liang, K.-Y., and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.

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

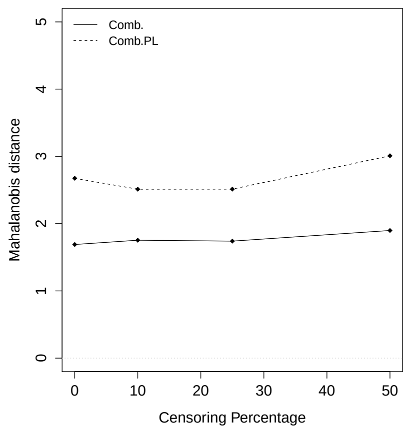
Achmad Efendi[1] **Geert Molenberghs**[2,1] **Samuel Iddi**[1]

[1] *I-BioStat, Katholieke Universiteit Leuven, Belgium*

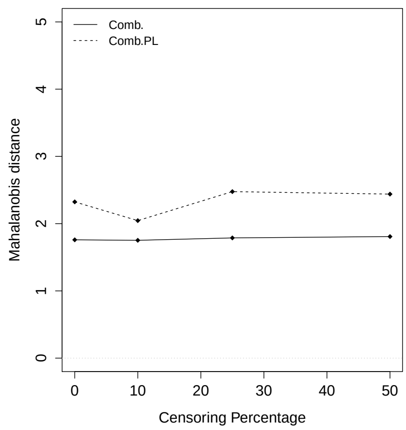
[2] *I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium*

Supplementary Materials

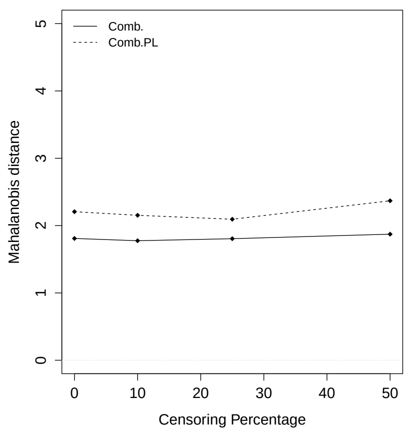
A Simulation Results



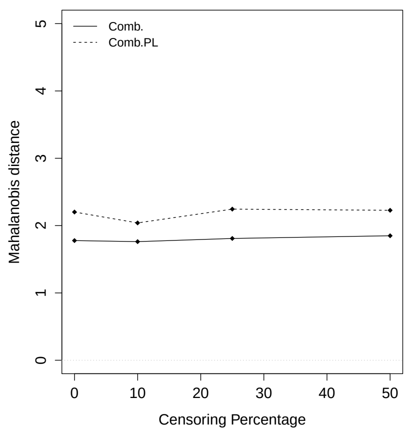
(a) 20 subjects



(b) 40 subjects

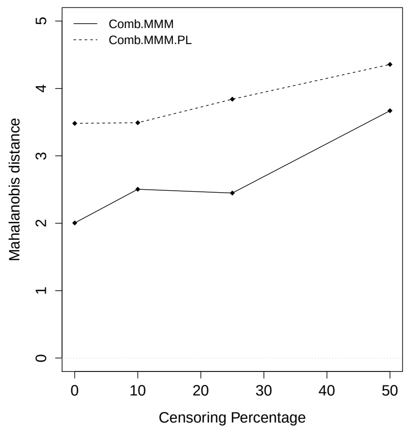


(c) 60 subjects

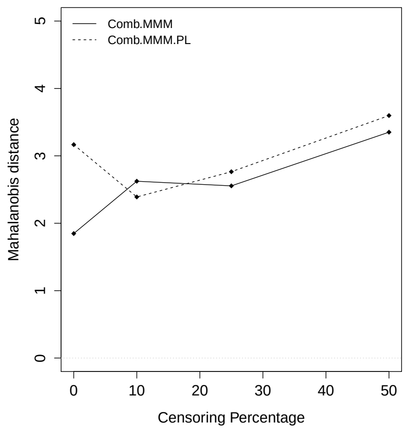


(d) 80 subjects

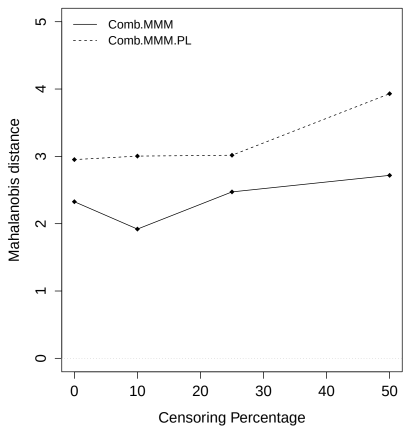
Figure A.1: *Mahalanobis distance for different sample size and censoring percentage*



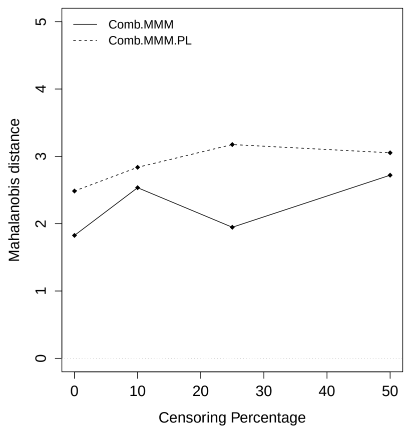
(a) 20 subjects



(b) 40 subjects



(c) 60 subjects



(d) 80 subjects

Figure A.2: *Mahalanobis distance for different sample size and censoring percentage for marginalized models*

Table A.1: *Simulation results. The combined model and the marginalized combined model with N=20, Full likelihood estimation.*

			WGN					WGN-M				
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	1.9973	1.0118	0.9601	0.0027	0.0014	1.9877	1.0112	0.9612	0.0096	0.0048
	λ	0.2	0.1991	0.0279	0.0273	0.0009	0.0046	0.2075	0.0300	0.0284	0.0084	0.0424
	α	2	2.2673	0.5935	0.6095	0.2673	0.1336	2.2762	0.5976	0.6078	0.0089	0.0039
	σ	0.25	0.2621	0.2302	0.1241	0.0120	0.0482	0.2618	0.2261	0.1241	0.0003	0.0010
% convergence					73.42					73.64		
10	ξ_1	2	2.0671	1.0402	1.0359	0.0671	0.0336	2.0611	1.0412	1.0406	0.0061	0.0029
	λ	0.2	0.1993	0.0295	0.0288	0.0007	0.0034	0.2086	0.0318	0.0302	0.0093	0.0468
	α	2	2.3655	0.8368	1.0250	0.3655	0.1828	2.3642	0.8373	1.0327	0.0014	0.0006
	σ	0.25	0.2747	0.2344	0.1244	0.0247	0.0987	0.2770	0.2217	0.1220	0.0023	0.0085
% convergence					70.72					70.32		
25	ξ_1	2	2.0379	1.0835	1.0497	0.0379	0.0189	2.0403	1.0828	1.0476	0.0024	0.0012
	λ	0.2	0.1975	0.0315	0.0304	0.0026	0.0127	0.2067	0.0337	0.0325	0.0093	0.0469
	α	2	2.6583	1.7714	1.9091	0.6583	0.3292	2.6083	1.5240	1.4658	0.0500	0.0188
	σ	0.25	0.2761	0.2314	0.1222	0.0261	0.1044	0.2737	0.2362	0.1245	0.0024	0.0088
% convergence					67.11					67.84		
50	ξ_1	2	2.1097	1.2593	1.1990	0.1097	0.0549	2.1129	1.2580	1.2013	0.0032	0.0015
	λ	0.2	0.2041	0.0395	0.0347	0.0041	0.0207	0.2163	0.0426	0.0377	0.0121	0.0593
	α	2	3.8012	14.9733	5.7834	1.8012	0.9006	3.8481	15.1684	5.8047	0.0469	0.0123
	σ	0.25	0.2998	0.3007	0.1447	0.0498	0.1993	0.3018	0.2978	0.1427	0.0019	0.0065
% convergence					59.67					59.17		

Table A.2: *Simulation results. The combined model and the marginalized combined model with N=20, Pairwise likelihood estimation.*

			WGN					WGN-M				
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	2.0061	0.9939	0.9611	0.0061	0.0031	1.9866	0.9931	0.9666	0.0196	0.0098
	λ	0.2	0.1973	0.0275	0.0271	0.0027	0.0134	0.2079	0.0303	0.0287	0.0106	0.0535
	α	2	2.3374	0.5694	0.6684	0.3374	0.1687	2.3342	0.5824	0.6587	0.0032	0.0014
	σ	0.25	0.2790	0.2008	0.1386	0.0290	0.1161	0.2692	0.2815	0.1458	0.0099	0.0354
% convergence					72.57					75.19		
10	ξ_1	2	2.0552	1.0153	1.0339	0.0552	0.0276	2.0653	1.0175	1.0168	0.0102	0.0049
	λ	0.2	0.1987	0.0292	0.0288	0.0013	0.0066	0.2086	0.0320	0.0306	0.0100	0.0501
	α	2	2.4078	0.7957	0.9836	0.4078	0.2039	2.3835	0.7818	0.9731	0.0243	0.0101
	σ	0.25	0.2677	0.2011	0.1425	0.0177	0.0709	0.2678	0.2074	0.1435	0.0001	0.0003
% convergence					74.74					74.63		
25	ξ_1	2	2.0356	1.0470	1.0571	0.0356	0.0178	2.0308	1.0488	1.0569	0.0048	0.0024
	λ	0.2	0.1982	0.0314	0.0305	0.0018	0.0089	0.2068	0.0337	0.0328	0.0085	0.0431
	α	2	2.7170	1.9355	2.1193	0.7170	0.3585	2.7079	1.9091	2.1047	0.0090	0.0033
	σ	0.25	0.2586	0.2144	0.1468	0.0086	0.0343	0.2536	0.2136	0.1490	0.0049	0.0191
% convergence					71.12					72.15		
50	ξ_1	2	2.1303	1.2336	1.1879	0.1303	0.0651	2.1306	1.2354	1.1747	0.0004	0.0002
	λ	0.2	0.2032	0.0392	0.0352	0.0032	0.0162	0.2143	0.0420	0.0384	0.0111	0.0546
	α	2	3.6153	11.3361	4.8598	1.6153	0.8076	3.6605	10.9005	4.8266	0.0452	0.0125
	σ	0.25	0.2762	0.2443	0.1586	0.0262	0.1046	0.2706	0.2366	0.1621	0.0056	0.0202
% convergence					63.53					64.60		

Table A.3: *Simulation results. The combined model and the marginalized combined model with N=40, Full likelihood estimation.*

			WGN					WGN-M				
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	1.9943	0.7185	0.6984	0.0057	0.0028	1.9966	0.7182	0.6990	0.0023	0.0012
	λ	0.2	0.1999	0.0196	0.0187	0.0001	0.0007	0.2075	0.0210	0.0203	0.0076	0.0382
	α	2	2.1060	0.3553	0.3792	0.1060	0.0530	2.1067	0.3553	0.3778	0.0007	0.0003
	σ	0.25	0.2575	0.1502	0.1018	0.0074	0.0298	0.2552	0.1551	0.1046	0.0023	0.0088
% convergence					84.32					85.32		
10	ξ_1	2	1.9968	0.7318	0.6996	0.0032	0.0016	1.9993	0.7313	0.7016	0.0025	0.0013
	λ	0.2	0.1987	0.0203	0.0204	0.0013	0.0067	0.2062	0.0215	0.0211	0.0075	0.0377
	α	2	2.1706	0.4641	0.5055	0.1706	0.0853	2.1688	0.4631	0.5059	0.0018	0.0008
	σ	0.25	0.2486	0.1577	0.1013	0.0014	0.0055	0.2461	0.1590	0.1046	0.0025	0.0102
% convergence					80.91					82.10		
25	ξ_1	2	1.9765	0.7655	0.7519	0.0235	0.0117	1.9795	0.7651	0.7564	0.0030	0.0015
	λ	0.2	0.2008	0.0224	0.0238	0.0008	0.0038	0.2088	0.0236	0.0250	0.0081	0.0401
	α	2	2.2665	0.7167	0.7417	0.2665	0.1333	2.2622	0.7142	0.7455	0.0044	0.0019
	σ	0.25	0.2574	0.1663	0.1044	0.0074	0.0297	0.2537	0.1685	0.1086	0.0038	0.0146
% convergence					78.49					79.74		
50	ξ_1	2	2.0210	0.8655	0.8579	0.0210	0.0105	2.0073	0.8657	0.8549	0.0137	0.0068
	λ	0.2	0.1995	0.0265	0.0249	0.0005	0.0023	0.2082	0.0279	0.0264	0.0086	0.0432
	α	2	3.1267	4.4612	3.2731	1.1267	0.5634	3.1353	4.4796	3.2758	0.0086	0.0027
	σ	0.25	0.2658	0.2156	0.1220	0.0158	0.0633	0.2633	0.2292	0.1235	0.0025	0.0094
% convergence					76.22					76.92		

Table A.4: *Simulation results. The combined model and the marginalized combined model with N=40, Pairwise likelihood estimation.*

					WGN					WGN-M		
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	1.9999	0.7119	0.6986	0.0001	0.0001	2.0009	0.7151	0.6926	0.0010	0.0005
	λ	0.2	0.1995	0.0194	0.0188	0.0005	0.0027	0.2077	0.0211	0.0201	0.0082	0.0411
	α	2	2.1189	0.3420	0.3862	0.1189	0.0594	2.1155	0.3403	0.3831	0.0034	0.0016
	σ	0.25	0.2533	0.1578	0.1242	0.0033	0.0132	0.2557	0.1566	0.1226	0.0024	0.0094
% convergence					84.89					84.32		
10	ξ_1	2	2.0117	0.7211	0.6925	0.0117	0.0059	2.0016	0.7213	0.7023	0.0101	0.0050
	λ	0.2	0.1984	0.0204	0.0202	0.0016	0.0081	0.2069	0.0219	0.0209	0.0085	0.0430
	α	2	2.1915	0.4676	.5126	0.1915	0.0957	2.1823	0.4664	0.5144	0.0092	0.0042
	σ	0.25	0.2542	0.1555	0.1140	0.0042	0.0169	0.2515	0.1657	0.1165	0.0027	0.0108
% convergence					81.17					81.97		
25	ξ_1	2	1.9492	0.7516	0.7326	0.0508	0.0254	1.9575	0.7536	0.7434	0.0083	0.0043
	λ	0.2	0.2009	0.0223	0.0232	0.0009	0.0044	0.2089	0.0238	0.0246	0.0080	0.0397
	α	2	2.2589	0.6978	0.7260	0.2589	0.1294	2.2699	0.7082	0.7445	0.0110	0.0049
	σ	0.25	0.2447	0.1602	0.1263	0.0053	0.0212	0.2491	0.1621	0.1227	0.0044	0.0181
% convergence					83.33					81.97		
50	ξ_1	2	2.0202	0.8564	0.8603	0.0202	0.0101	2.0179	0.8546	0.8607	0.0023	0.0011
	λ	0.2	0.2010	0.0267	0.0254	0.0010	0.0049	0.2093	0.0279	0.0268	0.0083	0.0415
	α	2	3.0683	4.6202	3.4117	1.0683	0.5342	3.0393	4.5169	3.3723	0.0290	0.0094
	σ	0.25	0.2522	0.1691	0.1339	0.0022	0.0090	0.2504	0.1729	0.1354	0.0019	0.0075
% convergence					80.52					81.04		

Table A.5: *Simulation results. The combined model and the marginalized combined model with N=60, Full likelihood estimation.*

			WGN					WGN-M				
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	2.0015	0.5856	0.5839	0.0015	0.0007	1.9954	0.5856	0.5791	0.0061	0.0030
	λ	0.2	0.1998	0.0159	0.0153	0.0002	0.0008	0.2067	0.0168	0.0162	0.0069	0.0343
	α	2	2.0661	0.2774	0.2856	0.0661	0.0330	2.0665	0.2776	0.2866	0.0005	0.0002
	σ	0.25	0.2486	0.1212	0.0890	0.0014	0.0056	0.2452	0.1293	0.0926	0.0034	0.0137
% convergence					90.42					91.41		
10	ξ_1	2	1.9960	0.5967	0.5957	0.0040	0.0020	1.9992	0.5970	0.6048	0.0032	0.0016
	λ	0.2	0.1991	0.0166	0.0167	0.0009	0.0047	0.2060	0.0175	0.0177	0.0069	0.0348
	α	2	2.1110	0.3542	0.3745	0.1110	0.0555	2.1098	0.3538	0.3775	0.0012	0.0006
	σ	0.25	0.2469	0.1242	0.0898	0.0031	0.0123	0.2446	0.1293	0.0930	0.0023	0.0094
% convergence					89.29					90.09		
25	ξ_1	2	2.0156	0.6230	0.6234	0.0156	0.0078	2.0031	0.6229	0.6252	0.0125	0.0062
	λ	0.2	0.1996	0.0180	0.0179	0.0004	0.0019	0.2066	0.0189	0.0189	0.0070	0.0351
	α	2	2.1579	0.5193	0.5667	0.1579	0.0790	2.1619	0.5221	0.5818	0.0040	0.0018
	σ	0.25	0.2420	0.1474	0.0969	0.0080	0.0322	0.2420	0.1422	0.0969	0.0000	0.0000
% convergence					86.06					86.06		
50	ξ_1	2	2.0186	0.7085	0.7017	0.0186	0.0093	2.0181	0.7083	0.7049	0.0005	0.0002
	λ	0.2	0.1992	0.0215	0.0218	0.0008	0.0040	0.2069	0.0225	0.0227	0.0078	0.0389
	α	2	2.5586	1.9445	1.9516	0.5586	0.2793	2.5689	1.9551	1.9535	0.0102	0.0040
	σ	0.25	0.2522	0.1771	0.1128	0.0022	0.0087	0.2509	0.1788	0.1135	0.0013	0.0051
% convergence					81.57					81.97		

Table A.6: *Simulation results. The combined model and the marginalized combined model with N=60, Pairwise likelihood estimation.*

					WGN					WGN-M		
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	1.9914	0.5836	0.5755	0.0086	0.0043	1.9948	0.5837	0.5714	0.0034	0.0017
	λ	0.2	0.1995	0.0158	0.0155	0.0005	0.0026	0.2072	0.0171	0.0166	0.0077	0.0385
	α	2	2.0786	0.2737	0.2949	0.0786	0.0393	2.0800	0.2744	0.2958	0.0014	0.0007
	σ	0.25	0.2461	0.1407	0.1166	0.0039	0.0155	0.2473	0.1403	0.1154	0.0012	0.0047
% convergence					87.72					87.41		
10	ξ_1	2	1.9844	0.5946	0.6053	0.0156	0.0078	1.9850	0.5942	0.6029	0.0006	0.0003
	λ	0.2	0.1989	0.0165	0.0172	0.0011	0.0055	0.2063	0.0176	0.0180	0.0074	0.0373
	α	2	2.1171	0.3557	0.3741	0.1171	0.0586	2.1095	0.3533	0.3757	0.0076	0.0036
	σ	0.25	0.2455	0.1285	0.1067	0.0045	0.0180	0.2419	0.1266	0.1111	0.0036	0.0145
% convergence					87.72					89.45		
25	ξ_1	2	2.0108	0.6172	0.6193	0.0108	0.0054	2.0074	0.6170	0.6258	0.0035	0.0017
	λ	0.2	0.1999	0.0181	0.0181	0.0001	0.0006	0.2067	0.0189	0.0189	0.0069	0.0344
	α	2	2.1601	0.5265	0.6197	0.1601	0.0800	2.1576	0.5244	0.6134	0.0025	0.0011
	σ	0.25	0.2372	0.1500	0.1092	0.0128	0.0513	0.2333	0.1492	0.1119	0.0039	0.0163
% convergence					87.87					89.13		
50	ξ_1	2	2.0083	0.7070	0.7107	0.0083	0.0041	1.9930	0.7058	0.7084	0.0153	0.0076
	λ	0.2	0.1993	0.0215	0.0220	0.0007	0.0033	0.2070	0.0224	0.0227	0.0076	0.0384
	α	2	2.5884	1.8619	1.7947	0.5884	0.2942	2.5426	1.7427	1.6735	0.0459	0.0177
	σ	0.25	0.2425	0.1519	0.1227	0.0075	0.0302	0.2345	0.1575	0.1279	0.0080	0.0328
% convergence					84.32					86.96		

Table A.7: *Simulation results. The combined model and the marginalized combined model with N=80, Full likelihood estimation.*

			WGN					WGN-M				
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	2.0010	0.5078	0.4875	0.0010	0.0005	2.0004	0.5077	0.4907	0.0006	0.0003
	λ	0.2	0.2000	0.0136	0.0134	0.0000	0.0002	0.2065	0.0144	0.0138	0.0066	0.0328
	α	2	2.0452	0.2345	0.2446	0.0452	0.0226	2.0470	0.2349	0.2432	0.0017	0.0009
	σ	0.25	0.2400	0.1064	0.0827	0.0100	0.0399	0.2410	0.1045	0.0813	0.0010	0.0041
% convergence					93.81					93.46		
10	ξ_1	2	2.0143	0.5156	0.5048	0.0143	0.0071	2.0126	0.5155	0.5054	0.0017	0.0008
	λ	0.2	0.1991	0.0143	0.0137	0.0009	0.0043	0.2056	0.0150	0.0142	0.0064	0.0322
	α	2	2.0840	0.2971	0.2884	0.0840	0.0420	2.0843	0.2969	0.2874	0.0003	0.0002
	σ	0.25	0.2398	0.1153	0.0852	0.0102	0.0408	0.2373	0.1152	0.0884	0.0025	0.0103
% convergence					92.94					93.81		
25	ξ_1	2	2.0195	0.5389	0.5383	0.0195	0.0098	2.0131	0.5389	0.5393	0.0064	0.0032
	λ	0.2	0.2001	0.0155	0.0159	0.0001	0.0005	0.2060	0.0161	0.0163	0.0059	0.0297
	α	2	2.1177	0.4270	0.4515	0.1177	0.0589	2.1232	0.4291	0.4527	0.0055	0.0026
	σ	0.25	0.2284	0.1308	0.0901	0.0216	0.0866	0.2309	0.1282	0.0873	0.0025	0.0111
% convergence					90.25					89.29		
50	ξ_1	2	2.0389	0.6112	0.6097	0.0389	0.0194	2.0249	0.6106	0.6115	0.0140	0.0069
	λ	0.2	0.2002	0.0186	0.0181	0.0002	0.0012	0.2070	0.0193	0.0188	0.0068	0.0337
	α	2	2.5180	1.9344	2.4326	0.5180	0.2590	2.5312	1.9465	2.4326	0.0132	0.0052
	σ	0.25	0.2406	0.1513	0.1056	0.0094	0.0374	0.2399	0.1515	0.1063	0.0008	0.0033
% convergence					87.26					87.41		

Table A.8: *Simulation results. The combined model and the marginalized combined model with N=80, Pairwise likelihood estimation.*

					WGN					WGN-M		
Cens.	Par.	True	Est.	s.e.	s.e.MC	Bias	Rbias	Est.	s.e.	s.e.MC	Bias	Rbias
00	ξ_1	2	2.0027	0.5089	0.5035	0.0027	0.0014	2.0082	0.5090	0.5022	0.0055	0.0028
	λ	0.2	0.1997	0.0136	0.0134	0.0003	0.0016	0.2064	0.0145	0.0141	0.0068	0.0338
	α	2	2.0561	0.2344	0.2495	0.0561	0.0280	2.0561	0.2347	0.2489	0.0000	0.0000
	σ	0.25	0.2402	0.1157	0.1014	0.0098	0.0392	0.2432	0.1200	0.0981	0.0030	0.0124
% convergence					90.09					89.13		
10	ξ_1	2	2.0195	0.5167	0.5046	0.0195	0.0097	2.0091	0.5157	0.5084	0.0104	0.0052
	λ	0.2	0.1992	0.0144	0.0141	0.0008	0.0041	0.2059	0.0151	0.0144	0.0068	0.0339
	α	2	2.0905	0.2995	0.2963	0.0905	0.0452	2.0935	0.3002	0.2961	0.0031	0.0015
	σ	0.25	0.2441	0.1229	0.0997	0.0059	0.0238	0.2413	0.1194	0.1026	0.0027	0.0111
% convergence					90.42					92.08		
25	ξ_1	2	2.0366	0.5343	0.5510	0.0366	0.0183	2.0310	0.5342	0.5441	0.0056	0.0027
	λ	0.2	0.2000	0.0155	0.0158	0.0000	0.0001	0.2060	0.0162	0.0162	0.0060	0.0300
	α	2	2.1161	0.4294	0.4616	0.1161	0.0581	2.1167	0.4305	0.4618	0.0006	0.0003
	σ	0.25	0.2220	0.1290	0.1037	0.0280	0.1121	0.2235	0.1427	0.1017	0.0015	0.0068
% convergence					90.74					90.42		
50	ξ_1	2	2.0252	0.6046	0.6041	0.0252	0.0126	2.0278	0.6043	0.6080	0.0025	0.0012
	λ	0.2	0.2000	0.0187	0.0182	0.0000	0.0001	0.2072	0.0193	0.0191	0.0072	0.0360
	α	2	2.4786	1.5215	1.7826	0.4786	0.2393	2.4789	1.5198	1.7800	0.0003	0.0001
	σ	0.25	0.2344	0.1459	0.1143	0.0156	0.0626	0.2336	0.1458	0.1151	0.0008	0.0032
% convergence					88.50					88.97		

B Software Code

```
/**************************************************************************
OBJECTIVE:  To analyze the Asthma dataset using the Weibull model with Gamma
frailty and random normal effects and its marginalized form.
DATASET: Example 9 of Duchateau & Janssen (2008);
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.
****************************************************************************/

proc sort data=asma1; by Patid;run;

*Weibull-Gamma-Normal - right censoring;
proc nlmixed data=asma1 tech=quanew qpoints=50 maxit=1000;
bounds lambda>0, alpha>0;
parms  Beta_1=-0.08 lambda=1 alpha=3.3 sigma1=1 sigma2=1;
rho=1;
eta = (Beta_1+b2)*(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 b2 ~ normal([0,0],[sigma1**2,0,sigma2**2]) subject=Patid;
estimate 'Var of R.E1.s' sigma1**2;
estimate 'Var of R.E2.s' sigma2**2;
run;

*Weibull-Gamma-Normal - right censoring - MMM;
proc nlmixed data=asma1 tech=quanew qpoints=50 maxit=1000;
bounds lambda > 0, alpha > 0;
parms  Beta_1=-0.08 lambda=1 alpha=3.3 sigma1=1 sigma2=1;
rho=1;
eta = (Beta_1+b2)*(Drug=1);
delta=eta-(sigma1*sigma1+sigma2*sigma2*(Drug=1)*(Drug=1))/2;
etas=delta+b1;
expeta = exp(etas);
c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
c1 = log(lambda) + log(rho) +
                          (alpha+1)*log(alpha)+ (rho-1)*log(Time) + etas
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
```

```
ll  = (status=0)*log(c0) + (status=1)*c1;
model Time ~ general(ll);
random b1 b2 ~ normal([0,0],[sigma1**2,0,sigma2**2]) subject=Patid;
estimate 'Var of R.E1.s' sigma1**2;
estimate 'Var of R.E2.s' sigma2**2;
run;
```