

A multilevel model for hierarchical, repeated, and overdispersed  
time-to-event outcomes and its estimation strategies

Peer-reviewed author version

EFENDI, Achmad & MOLENBERGHS, Geert (2013) A multilevel model for hierarchical, repeated, and overdispersed time-to-event outcomes and its estimation strategies. In: Journal of Biopharmaceutical Statistics, 23 (6), p. 1420-1434.

DOI: 10.1080/10543406.2013.834914

Handle: <http://hdl.handle.net/1942/14819>

# A Multilevel Model for Hierarchical, Repeated, and Overdispersed Time-to-event Outcomes and Its Estimation Strategies

Achmad Efendi<sup>1</sup>    Geert Molenberghs<sup>1,2</sup>

<sup>1</sup> *I-Biostat, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium*

<sup>2</sup> *I-Biostat, Universiteit Hasselt, Belgium*

## Abstract

The aim of this paper is to propose a multilevel combined model for repeated, hierarchical, and overdispersed time-to-event outcomes, extending the so-called combined model proposed by Molenberghs *et al* (2010), and using three different estimation strategies: full likelihood, pseudo-likelihood, and Bayesian estimation. For the first two estimation methods, we implemented the Alternating Imputation Posterior (AIP) algorithm (Clayton and Rasbash, 1999). It is shown that the multilevel combined model can be fitted nicely using all three estimation methods. In addition, the multilevel combined model has the advantage that it not only can capture the hierarchical structure of the data but also can accommodate overdispersion within the dataset. From our simulation results, it follows that the multilevel combined model performs well in terms of point estimation and its precision, fitted with the three different estimation methods. We also observed that pairwise likelihood estimation, a particular form of pseudo-likelihood, is more time-intensive than full likelihood and Bayesian estimation. However, pseudo-likelihood estimation is less sensitive to starting values.

**Some Keywords:** Bayesian estimation; Combined model; Maximum likelihood; Multilevel model; Pairwise likelihood; Random-effects model; Weibull distribution.

## 1 Introduction

The substantial growth in use of time-to-event outcomes started when Cox's proportional hazard model (Cox, 1972) was introduced; it has become the most popular model used in this context. While popular, it is not without limitations and hence a variety of extensions have been proposed. One of its basic assumption is independence between the survival times. In case there is possible association induced by multiple observations on the same patient, or in general on the same unit, the model tends to incorrectly estimate precision.

Building upon standard generalized linear models (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), there have been extensions towards accommodating the hierarchical structure for repeated survival outcomes via frailties and/or random effects. A regression model with frailty was fitted for repeated measures of recurrent times in McGilchrist and Aisbett (1991). Sargent (1998) introduced a general framework for random-effects survival data within the Cox proportional hazard tradition, the so-called hierarchical Cox model. O'Quigley and Stare (2002) discussed the concepts

underlying the development of frailties and random-effects models in survival. A nice frailty-model summary is offered in Duchateau and Janssen (2008). Simultaneously, Hinde and Demétrio (1998) addressed overdispersion. Building upon these strands of research, Molenberghs *et al* (2010) defined a broad modeling framework for exponential models to accommodate overdispersion and correlation in repeated, overdispersed binary, count, and time-to-event outcomes. The authors combined so-called conjugate random effects for overdispersion with generalized linear mixed model concepts (GLMM, Engel and Keen (1994), Breslow and Clayton (1993), Wolfinger and O'Connell (1993)) for between-subject effects. For simplicity, the model is merely called combined model. A combined survival-type model is presented in Molenberghs *et al* (2011), where also censorship is allowed for. These authors also supplemented the model-fitting tool box with a method based on pairwise likelihood.

Research on hierarchical or, equivalently, multilevel models is gaining popularity, whether in the survival-data setting or beyond. Such models allow for both longitudinal as well as clustered aspects in the data. This broad umbrella makes the model versatile and widely applicable. For instance, in clinical trials, patients often come from various centers, hospitals, regions, or countries, and/or are followed up repeatedly over time. Yau (2001) described a method for modeling survival data with random effects following the approach by McGilchrist (1993) and generalized the estimation procedure to models with arbitrary number of levels. Wang *et al* (2007) incorporated two random components in the survival frailty model to simultaneously account for heterogeneity between centers and for the within-patient autoregressive structure. In terms of the frailty distribution, recently Kim and Dey (2008), rather than using gamma-distributed frailties, developed a class of heavy-tailed distributions. They fitted frailty models to multilevel survival data and employed Bayesian estimation. Overall, the aforementioned multilevel models mentioned so far accommodate an individual or a cluster level effect through frailties or random effects. Our goal is to formulate and fit a multilevel survival model by following the approach of Molenberghs *et al* (2011), so as to accommodate the hierarchical data structure of the data through normal random effects and overdispersion through conjugate random effects, often also called frailties. We offer three estimation strategies: maximum likelihood, pairwise likelihood, and Bayesian estimation. Having several tools is useful, especially when different analyses produce rather different point and interval estimates (Browne and Draper, 2002).

The rest of the article is organized as follows. In Section 2, two motivating case studies are described, with their analyses reported in Section 5. The proposed combined model is described and further studied in Section 3. The estimation strategies will be presented in Section 4. Section 6 describes design and results of a simulation study.

## 2 Motivating Case Studies

### 2.1 Comet Assay Data

A comet assay, a common name for the single cell gel electrophoresis assay, is an easy-to-perform and sensitive technique for the detection of DNA damage at the level of the individual eukaryotic cell. Specifically, the data presented in this paper refer to four groups of six male rats that received a daily oral dose of a compound in three dose levels (low, medium, high) or vehicle control. 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 are 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 were 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. It is the percentage of DNA fragments 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.

## 2.2 Cardiovascular Health Study Data

These data encompass study endpoint events which occurred prior to July 1, 2002 and include all incident and recurrent MI, angina, CHF, stroke, TIA, and claudication events and all deaths through June 30, 2002. The study involves recurrent events with respect to a particular type of event. This is dealing with intra subject correlation; multiple events recorded per subject. Furthermore, there are eleven types of event: no event, TIA, MI, angioplasty, angina, coronary artery bypass surgery, stroke, other deaths (non-CHD), CHF, ECG MI (silent), claudication, and other CHD deaths. Additionally, there are 5772 patients or participants involved in the study, 3416 (59%) with a single observation from one event type. Among patients with more than one observation, 742 have two observations and 471 have three. Every event constitutes a record and participants without events are represented by means of single record spelling out that fact. Event-independent covariates, such as age, body mass index, hypertension, death, are repeated in every record for that particular patient.

## 3 Methodology

We briefly present and build upon the combined model of Molenberghs *et al* (2010) and Ghebretinsae *et al* (2011). Ample detail can be found there. The combined model brings together two sets of random effects. One set of random effects, usually normally distributed, accounts for the hierarchical structure of the data while the other, often conjugate random effects, is used for accommodating overdispersion. We build upon this concept for the context of survival data, in a hierarchical context, where the hierarchy is made up of repeated events and other types of clustering, simultaneously. Non-informative censoring is allowed for. It can take the form of right-, left-, and interval censoring. Details are given in Section 4.1. Precisely, the model is specified as:

$$Y_{ij} | \mathbf{b}_i, \theta_{ij} \sim \text{Weibull}(\rho, k_{ij}), \quad (1)$$

$$k_{ij} = \lambda \theta_{ij} e^{\mathbf{x}'_{ij} \boldsymbol{\xi} + \mathbf{z}'_{ij} \mathbf{b}_i}, \quad (2)$$

$$\mathbf{b}_i \sim N(0, D), \quad (3)$$

$$\theta_{ij} \sim \text{Gamma}(\alpha, \beta). \quad (4)$$

with  $Y_{ij}$  the time-to-event outcome of individual  $i$  and observation  $j$ . The model extends the general survival model with random effects in the Cox proportional hazard setting (Sargent, 1998), where  $\kappa_{ij}$  is the GLMM component with  $\mathbf{b}_i$  normally distributed random effects,  $\lambda$  the shape parameter of the Weibull distribution, and  $\rho$  its scale parameter. The conjugate random effects,  $\theta_{ij}$ , follow a Gamma distribution with parameters  $\alpha$  and  $\beta$ .

Up to here, the model allows for one hierarchical level only. This falls short of analyzing the comet assay data. For these, it is necessary to extend (1)–(4) by including an additional normal random

effect to account for sub-cluster (slide) with cluster (animal). The model becomes:

$$Y_{i_1 i_2 i_3} | \mathbf{b}_{i_1}, \mathbf{b}_{i_1 i_2}, \theta_{i_1 i_2 i_3} \sim \text{Weibull}(\rho, k_{i_1 i_2}), \quad (5)$$

$$k_{i_1 i_2 i_3} = \lambda \theta_{i_1 i_2 i_3} e^{\mathbf{x}'_{i_1 i_2 i_3} \boldsymbol{\xi} + \mathbf{z}'_{i_1 i_2} \mathbf{b}_{i_1} + \mathbf{z}'_{i_1 i_2 i_3} \mathbf{b}_{i_1 i_2}}, \quad (6)$$

$$\mathbf{b}_{i_1} \sim N(0, D_1), \quad (7)$$

$$\mathbf{b}_{i_1 i_2} \sim N(0, D_2), \quad (8)$$

$$\theta_{ijk} \sim \text{Gamma}(\alpha, \beta). \quad (9)$$

The extended notation is straightforward. Let us term this the Weibull-gamma-normal-normal model. To ensure identifiability, it is common to impose  $\alpha\beta \equiv 1$ , in both (4) and (9). The two-cluster-level model (5)–(9), is one of four possible hierarchical overdispersed non-Gaussian models, proposed in Ghebretinsae *et al* (2011) for comet data.

We can easily extend model (5)–(9) to a more complex combined model for  $m$  cluster levels, with additionally presence of overdispersion. Evidently, while such model is useful in practice and easy to write down, estimation will become increasingly complex with growing  $m$ . Therefore, we propose to make use of *Alternating Imputation Posterior* (AIP), introduced by Clayton and Rasbash (1999) and further studied by Ecochard and Clayton (2002). Suppose we have a general multilevel combined model with  $m$  cluster levels:

$$Y_{i_1 i_2 \dots i_m} | \mathbf{b}_{i_1}, \mathbf{b}_{i_1 i_2}, \dots, \mathbf{b}_{i_1 i_2 \dots i_{m-1}}, \theta_{i_1 i_2 \dots i_m} \sim \text{Weibull}(\rho, k_{i_1 i_2 \dots i_m}), \quad (10)$$

$$k_{i_1 i_2 \dots i_m} = \lambda \theta_{i_1 i_2 \dots i_m} \exp \left( \mathbf{x}'_{i_1 i_2 \dots i_m} \boldsymbol{\xi}_{i_1 i_2 \dots i_m} + \sum_{\ell=1}^m \mathbf{z}_{i_1 \dots i_\ell} \mathbf{b}_{i_1 \dots i_\ell} \right). \quad (11)$$

The  $\theta_{i_1 i_2 \dots i_m}$  are Gamma distributed, whereas all  $\mathbf{b}_{i_1 \dots i_\ell} \sim N(0, D_\ell)$ . Using the AIP algorithm, the model can be partitioned such that, for  $\ell = 1, \dots, m$ , the  $j$ th order random-effects structure is fitted by considering

$$o^{(j)} = \exp \left( \sum_{\ell \neq j} \mathbf{z}_{i_1 \dots i_\ell} \mathbf{b}_{i_1 \dots i_\ell} \right)$$

a known offset. At the  $t$ th iteration, the following steps are carried out, for each  $\ell = 1, \dots, m$ :

1. Fit the  $j$ th nested sub-model using the offset  $o^{(j)}$  calculated from current values of the random effects.
2. Sample the model parameters from an approximation to their joint posterior distribution  $[\boldsymbol{\xi} | \mathbf{y}, \mathbf{x}, o^{(j)}]$ .
3. Sample  $\mathbf{b}_{i_1 \dots i_j}$  from  $[\mathbf{b}_{i_1 \dots i_j} | \mathbf{y}, \mathbf{x}, o^{(j)}]$ , which is normally distributed.

The AIP has been implemented in both a sequential as well as a parallel fashion (Ecochard and Clayton, 2002). The overall estimate  $(\boldsymbol{\xi}, \mathbf{b})$  is the mean of the estimates at each step, whereas its variance is the mean of the variances plus the variance of the estimates, using the law of iterated expectations. Note that our method follows the principles laid out in (Ecochard and Clayton, 2002, page 327), using their so-called alternative proposal, where a Gaussian approximation is obtained using Rao-Blackwellization.

## 4 Estimation Strategies

Within each AIP set, the model needs to be fitted. This can be done using the three distinct strategies of full likelihood, pairwise likelihood, and Bayesian nature, respectively.

### 4.1 Maximum Likelihood Estimation

Within each AIP step, we fit a combined model with one normally distributed random effect as well as a conjugate random effect. A closed-form expression, useful for maximum likelihood estimation, is given in Molenberghs *et al* (2011), and using the partial marginalization of Molenberghs *et al* (2010). The idea behind this is to integrate the conditional density over the gamma random effects only, leaving the normal random effects in the density function. For the Weibull case, using two levels, this produces:

$$f(y_{ij}|\mathbf{b}_i) = \frac{\lambda \kappa_{ij} e^{\mu_{ij}} \rho y_{ij}^{\rho-1} \alpha_j \beta_j}{(1 + \lambda \kappa_{ij} e^{\mu_{ij}} \beta_j y_{ij}^{\rho})^{\alpha_j+1}}. \quad (12)$$

The same approach is followed when censoring occurs, simply by integrating (12) over the interval  $[0, C_{ij}]$  in the right-censored case:

$$f(C_{ij}|\mathbf{b}_i) = \int_{C_{ij}}^{+\infty} f(y_{ij}|\mathbf{b}_i) dy_{ij} = \frac{1}{(1 + \lambda \kappa_{ij} e^{\mu_{ij}} C_{ij}^{\rho})^{\alpha_j}}. \quad (13)$$

Left- and interval-censoring are handled analogously.

It is then straightforward to obtain the fully marginalized probability by numerically integrating the normal random effects out of (12) for the uncensored cases and (13) for the right-censored cases, using a common statistical tool, e.g., the SAS procedure NLMIXED.

### 4.2 Pairwise Likelihood Estimation

As an alternative to full likelihood, we 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. le Cessie and van Houwelingen (1991) and Geys, Molenberghs, and Ryan (1999) replace the proper 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 so-called pseudo-likelihood function. Grouping the outcomes for subject  $i$  into a vector  $\mathbf{Y}_i$ , the contribution of the  $i$ th 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 observations. Otherwise,  $p\ell_i = f(y_{i1})$  is to be used. Pairwise likelihood is a special case of pseudo-likelihood, studied in detail in Arnold and Strauss (1991) and Molenberghs and Verbeke (2005). When it comes to implementation, the ideas of maximum likelihood are used, but applied to the pairs rather than to the full set of measurements for a particular unit. This creates a problem for precision estimation, but using standard sandwich-estimator ideas, a consistent estimate of the variance-covariance matrix of the estimated parameters can easily be obtained Molenberghs and Verbeke (2005).

### 4.3 Bayesian Estimation

Unlike with maximum and pairwise likelihood, Bayesian estimation does not combine with AIP. Rather, direct implementation is possible using standard statistical programs such as Winbugs, R2Winbugs, and R2jags. These packages allow for Monte Carlo Markov Chain based integration over multiple random effects (Browne and Draper, 2002).

Consider fitting Weibull-gamma-normal-normal model (5)–(9). In the context of the comet assay data,  $Y_{i_1 i_2 i_3}$  represents tail intensity, in particular the  $i_3$ th ( $i_3 = 1, \dots, 50$ ) measurement in the  $i_2$ th slide ( $i_2 = 1, 2, 3$ ) from the  $i_1$ th animal ( $i_1 = 1, \dots, 27$ ). The model can be expressed in the form (5), with

$$k_{i_1 i_2 i_3} = \lambda \theta_{i_1 i_2 i_3} e^{\xi_1 x_{1i_1 i_2 i_3} + \xi_2 x_{2i_1 i_2 i_3} + \xi_3 x_{3i_1 i_2 i_3} + \xi_4 x_{4i_1 i_2 i_3} + b_{a,i_1} + b_{s,i_1 i_2}}, \quad (14)$$

where  $\theta_{i_1 i_2 i_3} \sim \text{Gamma}(\alpha, \beta)$ ,  $b_{a,i_1} \sim N(0, d_1)$  and  $b_{s,i_1 i_2} \sim N(0, d_2)$ .  $\theta_{i_1 i_2 i_3}$  is the conjugate random effect,  $b_{a,i_1}$  is the random intercept corresponding to the animal specific effect, and  $b_{s,i_1 i_2}$  is the random intercept that corresponds to the slide effect  $i_2$  within animal  $i_1$ . Furthermore,  $x_{1i_1 i_2 i_3}$ ,  $x_{2i_1 i_2 i_3}$ ,  $x_{3i_1 i_2 i_3}$ , and  $x_{4i_1 i_2 i_3}$  are indicators for the dose given to the animals.

## 5 Data Analysis

Both sets of data introduced in Section 2 will be analyzed using model (14), by means of all three estimation methods detailed in Section 4.

The overdispersion parameter  $\alpha$  is relatively difficult to estimate, due to the large number of 50 observations per slide, compared to only 3 slides per animal in the comet assay data. Another problem occurs with the CHS data, where the random-effects variance is hard to estimate in the presence of overdispersion. This can be explained because 3416 out of 5772 subjects, i.e., 59%, contribute a single observation only. For the same reason, it is hard to pinpoint overdispersion. These issues are brought forward to the simulation study, presented in the next section.

For the aforementioned reason, the overdispersion parameter is dropped from the model. The model does retain its two levels of clustering: animal and slide in the comet assay case, and patient and event type in the CHS study. Precisely, the model takes the form

$$k_{i_1 i_2 i_3} = e^{\beta_0 + \beta_1 x_{1i_1 i_2 i_3} + \beta_2 x_{2i_1 i_2 i_3} + \beta_3 x_{3i_1 i_2 i_3} + \beta_4 x_{4i_1 i_2 i_3} + b_{a,i_1} + b_{s,i_1 i_2}}$$

for the comet assay data, while for CHS we have:

$$k_{i_1 i_2 i_3} = e^{\beta_0 + \beta_1 G_{i_1 i_2 i_3} + \beta_2 BMI_{2i_1 i_2 i_3} + \beta_3 BMI_{3i_1 i_2 i_3} + \beta_4 H_{i_1 i_2 i_3} + b_{p,i_1} + b_{e,i_1 i_2}},$$

where  $G$ . refers to gender,  $BMI_2$ . indicates overweight,  $BMI_3$ . indicates obesity, and  $H$ . stands for hypertension. The model was fitted using the SAS procedure NLMIXED, supplemented with a SAS macro. Results can be found in Tables 1, and 2.

Comfortingly, all analyses indicate that there is generally little difference between the three estimation methods. For some parameters, e.g.,  $\beta_0$  and  $\sigma_e^2$  in the CHS study, the difference is slightly more pronounced. In terms of precision, maximum likelihood provides the smallest standard errors, and pairwise likelihood the largest. This is not unexpected, because pairwise likelihood, due to its non-likelihood basis, does not enjoy the same optimality properties as full maximum likelihood.

A graphical display to assess the fit of the model to the comet assay data is given in Figure 1. Precisely, the average observed and fitted profiles are plotted (tail intensity), with both the slide and

animal levels as basis of replication. The fitted values were generated based on equations (5) and (9), i.e., the multilevel model with two normal random effects. The steps involved are: fitting the multilevel model to then obtain parameter estimates and empirical Bayes estimates of one random effect, after which the other are generated using the estimated parameters in the last AIP iteration by way of offset. Thereafter, the predicted values of the Weibull distributed random variables are obtained, using the aforementioned estimates and given the covariates. Additionally, for each plot, we compare it to the fit obtained from a simpler model with only one normal random effect. As can be seen from the figure, the fitted values of the multilevel model exhibit slightly higher values that are more similar to the observed values than these obtained from the simpler model.

Furthermore, a formal assessment of the treatment effects in the comet assay case and of the covariates for CHS, based on every one of the three estimation strategies is presented in Table 3. As can be seen from that table, treatment-effect levels in the comet assay are similar in strength across estimation methods. Furthermore, although all covariates are statistically significant in the CHS study, we find that the covariates' importance is ordered across estimation methods, with relatively weaker evidence under pseudo-likelihood and strongest in the Bayesian case.

The contrast between the Bayesian and other strategies can be seen as an analysis of the sensitivity of the model to the specification of the prior. While we considered one choice here only, users can contrast Bayesian estimation under a variety of priors with the likelihood and pseudo-likelihood estimates.

Also, in line with Molenberghs *et al* (2010), our method can be seen as an assessment of fit relative to more conventional methods, where a model with either gamma or normal random effects would be considered, but not both. These features, together with the fact that the Weibull family is very flexible, owing to the inclusion of a shape parameter, makes the proposed model family very general, which has a beneficial impact on the quality of fit. Evidently, targeted goodness-of-fit tools could usefully be constructed, a topic of further research.

## 6 Simulation Study

Our main aim is to evaluate the performance of the multilevel combined model in terms of point and precision estimation, under full likelihood, pairwise likelihood, and in a Bayesian way. The simulation study was designed to range across different parameter and sample-size settings. For the AIP algorithm to work, each sub-model should properly allow fitting. Assessing the feasibility of this was therefore also included among the goals of the simulation study.

First, we generated dataset under the set of true parameters:  $\beta = 2$ ,  $\lambda = 0.2$ ,  $\alpha = 2$ , and  $\sigma = 0.5$ . The one covariate is taken to follow a  $N(0, 0.1)$  distribution.  $\lambda$  is the shape parameter of the Weibull distribution; we set the scale parameter  $\rho = 1$ , i.e., the exponential distribution.  $\alpha$  is used to generate gamma distributed frailties, with parameterization  $G(\alpha, 1/\alpha)$  (Duchateau and Janssen, 2008). Lastly, the random effects are generated from a  $N(0, \sigma^2)$ . We consider sample sizes of 20 and 40, with 10 repeated measures per subject. Combined with the three estimation methods, this makes for six settings, under each of which 500 datasets were generated. The simulation result are reported in Table 4.

The average of the parameter estimates (Est.), average of the estimated standard errors (s.e.), the standard error of the parameter estimates [s.e.(MC)], the bias, and the relative bias are reported. The relative bias ranges from 0.0005 to 0.4869. We found that among four parameters, the relative



bias gets bigger, with the smallest value for  $\lambda$ , then  $\beta_1$  and  $\alpha$ , and the largest one for  $\sigma$ . This behavior occurs under all three estimation methods. For the first setting, maximum likelihood and pairwise likelihood converged in roughly 72% of the cases; for the second setting this became a little over 80%. In the Bayesian case, in contrast, the convergence rate is virtually 100%. This behavior will be revisited in the ensuing simulation setting, for the multilevel combined model. While pairwise likelihood takes more time, it converges slightly more easily as the pairwise contributions are much stabler and easier to manipulate than the multivariate contributions needed under full likelihood. It also is considerably less sensitive to starting values. The precision estimates are adequate, as shows from the agreement between s.e. and s.e.(MC). Overall, the behavior of the combined model is satisfactory.

An additional simulation run was carried out with true parameters  $\beta = 2$ ,  $\lambda = 0.2$ ,  $\alpha = 2$ , and  $\sigma = 0.5$ , but with sample sizes of 5 and 10, and further with number of observations 2, 5, 10, 20, and 40. Details are omitted here. From comparing the results across sample sizes we understand better the difficulty in convergence for the case studies, when the overdispersion parameter is present. Precisely, we learn that, for sample size 5, the larger the number of observations, from 5, over 10 and 20, to 40, the less accurate the estimate of the overdispersion parameter  $\alpha$ . This result, at first sight counterintuitive, can be explained as follows. The overdispersion parameter in many models will exhibit partial or total aliasing with the intercept. Due to a lack of precision, this may not be noticeable in small samples, but the effect may become more visible with increasing sample sizes. This result should not be worrisome, because the estimation of all other model parameters, whether covariate effects, variance components, or shape parameters is unaffected by it.

Next, let us turn to the multilevel combined model. True parameters in this simulation are  $\beta = 2$ ,  $\lambda = 0.2$ ,  $\alpha = 2$ ,  $\sigma_1 = 0.5$ , and  $\sigma_2 = 0.5$ . The setting is largely similar to the setting for the combined model above, with now additional random effects. The assumptions about the covariate and the scale parameter are retained. Both random effects are generated from normal distribution  $N(0, \sigma_1^2)$  and  $N(0, \sigma_2^2)$ . We set the number of first-level clusters (subjects) to 20, and the second level to 20 and 40. The number of replicates is retained at 10. Again combining with the three estimation methods, we reach also here six settings. For each, 500 sets of data were generated. The simulation results are reported in Table 5.

From these results, it follows that, in all settings, we reach (virtually) 100% of convergences. All estimates, except for the variance components, come within close reach of their true values. There is good agreement across estimation strategies. This confirms the good point estimate performance of the multilevel combined model. We can also see that when the second-level sample size doubles, the relative bias of the fixed effect estimates halves. Next, turning to  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$ , we see that these variance components rather deviate from their true values.  $\hat{\sigma}_1^2$  is slightly underestimated by full likelihood and Bayesian estimation whereas using pairwise likelihood it is somewhat overestimated. As for  $\hat{\sigma}_2^2$ , the simulation results shows that model fitting using full likelihood and Bayesian estimation underestimates it, while pairwise likelihood produces some overestimation. Still, for the estimate, full likelihood and Bayesian estimation give similar relative bias whereas pairwise likelihood yields higher relative bias. Overall, it can be observed that increasing the sample size decreases the relative bias. Also, in terms of time consumed for model fitting, full likelihood shares time requirements with Bayesian estimation, i.e., around 22 minutes and 45 minutes, respectively, for sizes 20 and 40, respectively, using a VSC cluster computer, while pairwise likelihood needs considerably more time (3 and 6 hours on average) using the same computing tools. Lastly, we observed that the s.e. and s.e.(MC) reach almost similar values, in all settings via the three different estimation strategies, confirming the good precision performance of the multilevel combined model.

The contrast between the Bayesian and other strategies can be seen as an analysis of the sensitivity of the model to the specification of the prior. While we considered one choice here only, users can contrast Bayesian estimation under a variety of priors with the likelihood and pseudo-likelihood estimates.

## 7 Concluding Remarks

We have shown that, for time-to-event outcomes, the ordinary combined model introduced by Molenberghs *et al* (2010) can be extended to a multilevel combined model. The multilevel combined model not only captures multiple hierarchical structures in the data, but also allows for overdispersion. Our extended model poses numerical challenges, which we have overcome by the use of the AIP algorithm and by offering a triple of estimation methods.

Practically, model fitting is achieved by putting together the combined model as a sub-model and the AIP algorithm of Clayton and Rasbash (1999). It is shown, through targeted simulation, that the multilevel combined model proposed here can be fitted without too much difficulty. The model fitting can easily be done through common statistical software. For example, we combined the concept of partial marginalization (Molenberghs *et al*, 2010) to enable the use of the SAS procedure NLMIXED, surrounded by a dedicated SAS macro.

We have shown that several estimation strategies can be used to fit this model, adding flexibility to the modeler's toolkit. We focused on full likelihood, pairwise likelihood, and Bayesian estimation. The ordinary combined model was given attention as a dedicated sub-model of the one proposed. Broadly, in cases where the simpler model fits well, so does the multilevel version. This result, derived from simulations, is useful in that it suggests that the extended model does not add undue additional computational complexity.

It has been shown that the multilevel combined model performs well in terms of point estimates and corresponding precision, with the three different estimation methods. Additionally, we observed that pairwise likelihood estimation needs notably more time to fit the model as compared to full likelihood and Bayesian estimation. This seems a drawback but is offset somewhat by the fact that the pairwise likelihood version is considerably less sensitive to starting values. Finally, in the case of larger number of cluster levels, i.e., more than two, we expect that it will still be straightforward to fit the multilevel combined model via the three different estimation strategies. Additional research may be in place to further optimize computation, when more hierarchical levels are needed than in the case studies considered here.

## Acknowledgment

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

## References

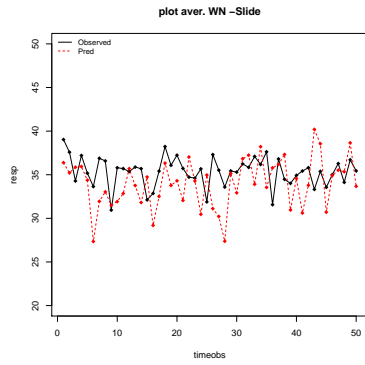
Arnold, B.C. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Sankhya B*, **53**, 233–243.

- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Browne, W.J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, **1**, 473–514.
- Clayton, D. and Rasbash, J. (1999). Estimation in large crossed random-effect models by data augmentation. *Journal of the Royal Statistical Society, Series A* **162**, 425–436.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. New York: Springer.
- Ecochard, R. and Clayton, D.J. (2002). Fitting complex random effect models with standard software using data augmentation: application to a study of male and female fecundability. *Statistical Modelling*, **1**, 319–331.
- 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.
- Ghebretinsae, A.H., Faes, C., Molenberghs, G., De Boeck, M., and Geys, H. (2011). A Bayesian, generalized frailty model for comet assays. *Submitted for publication*.
- Hinde, J. and Demétrio, C.G.B. (1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.
- Kim, S. and Dey, D.K. (2008). Modeling multilevel survival data using frailty models. *Communications in Statistics, Theory and Methods*, **37**, 1734–1741.
- 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.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. London: Chapman & Hall/CRC.
- McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics*, **49**, 221–225.
- McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461–466.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Molenberghs, G., Verbeke, G., Demétrio 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. (2011). A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *Submitted for publication*.

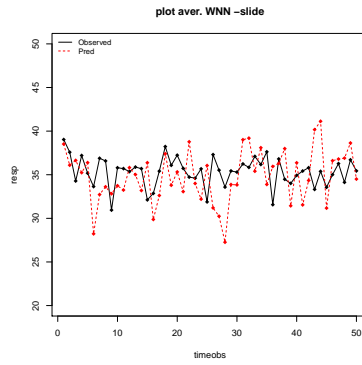
- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series B*, **135**, 370–384.
- O'Quigley, J. and Stare, J. Proportional hazards models with frailties and random effects. *Statistics in Medicine*, **21**, 3219–3233.
- Renard, D., Molenberghs, G., and Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics and Data Analysis*, **44**, 649–667.
- Sargent, D.J. (1998). A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics*, **54**, 1486–1497.
- Wang, K., Yau, K.K.W., Lee, A. H., and Mclachlan, G.J. (2007). Multilevel survival modeling of recurrent urinary tract infections. *Computer Methods and Programs in Biomedicine*, **87**, 225–229.
- Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233–243.
- Yau, K.K.W. (2001). Multilevel models for survival analysis with random effects. *Biometrics* **57**, 96–102.

**Table 1:** Comet study. Weibull model with two normal random effects. (e.c.s.e.: empirically corrected s.e.)

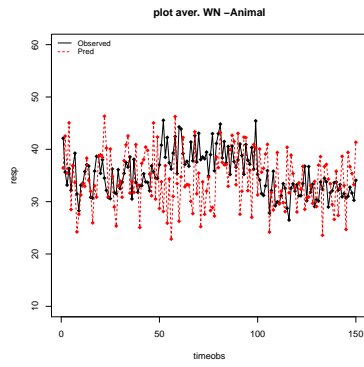
	Bayesian	Full likelihood	Pseudo-likelihood
Parameter	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.; e.c.s.e.)
$\xi_0$	-3.7940 (1.5800)	-2.4712 (0.0773)	-2.4087 (0.0083; 0.5621)
$\xi_1$	-1.7950 (0.1321)	-2.8142 (0.0908)	-2.7642 (0.0109; 0.3770)
$\xi_2$	-1.0220 (0.1301)	-3.0400 (0.0917)	-3.1243 (0.0112; 0.4152)
$\xi_3$	-1.2640 (0.1313)	-3.2871 (0.0927)	-3.1833 (0.0116; 0.4492)
$\xi_4$	-1.4890 (0.1314)	-1.7894 (0.1075)	-1.7878 (0.0106; 0.2438)
$\gamma$	2.0370 (0.2910)	1.4173 (0.0189)	1.3907 (0.0030; 0.2316)
$\sigma_a$	0.1092 (0.0352)	0.1629 (0.0276)	0.1663 (0.0064; 0.0566)
$\sigma_s$	0.2198 (0.0282)	0.2192 (0.0248)	0.1370 (0.0074; 0.0566)



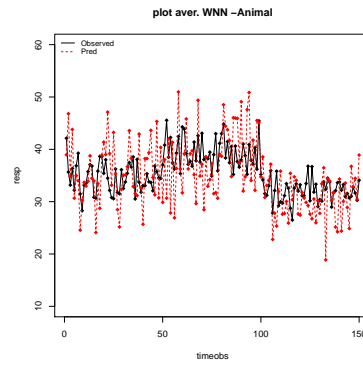
(a) model with 1 RE



(b) model with 2 RE



(c) model with 1 RE



(d) model with 2 RE

**Figure 1:** *Comet assay data:* Plots of average profiles based on subject levels of slide and animal. 'WN' refers to Weibull-normal model and 'WNN' to Weibull-normal-normal model.

**Table 2:** CHS study. Weibull model with two normal random effects. (e.c.s.e.: empirically corrected s.e.)

	Bayesian	Full likelihood	Pairwise likelihood
Parameter	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.; e.c.s.e.)
<i>without censoring</i>			
$\xi_0$	-9.711 (0.3586)	-17.530 (0.16125)	-25.698 (0.16282; 1.40839)
$\xi_1$	0.449 (0.0496)	0.1030 (0.01753)	0.4261 (0.02670; 0.19541)
$\xi_2$	-0.249 (0.0535)	-0.1890 (0.01964)	-0.1564 (0.02956; 0.19833)
$\xi_3$	-0.214 (0.0658)	-0.1925 (0.02358)	-0.1689 (0.03687; 0.14530)
$\xi_4$	0.281 (0.0503)	0.0927 (0.01753)	0.2916 (0.02678; 0.16527)
$\gamma$	3.787 (1.5400)	2.0285 (0.01495)	3.1846 (0.02002; 0.17481)
$\sigma_p^2$	1.529 (0.0492)	0.1081 (0.00512)	4.7469 (0.44571; 3.38655)
$\sigma_e^2$	18.28 (4.4340)	1.6896 (0.11924)	1.8482 (0.15347; 1.44623)
<i>with censoring</i>			
$\xi_0$	-19.4095 (0.0940)	-36.3169 (0.37887)	-42.3589 (0.32284; 1.80329)
$\xi_1$	0.0082 (0.0596)	0.1096 (0.02272)	0.0451 (0.01874; 0.03453)
$\xi_2$	-0.3220 (0.0994)	-0.0110 (0.02530)	-0.0849 (0.02175; 0.11336)
$\xi_3$	-0.3524 (0.1204)	-0.0621 (0.03031)	-0.1059 (0.02580; 0.09152)
$\xi_4$	0.1932 (0.0572)	0.1045 (0.02252)	0.1156 (0.01875; 0.05628)
$\gamma$	14.512 (2.4006)	4.2893 (0.03967)	5.0123 (0.03814; 0.27125)
$\sigma_p^2$	5.7056 (0.4133)	0.3341 (0.04805)	0.0403 (0.00068; 0.00107)
$\sigma_e^2$	17.345 (8.9886)	0.1506 (0.00291)	0.1479 (0.01800; 0.02427)

**Table 3:** Comet assay and CHS study. Wald test for covariate effect's assessment.

Covariate	Bayesian		Full likelihood		Pseudo- likelihood	
	Z-value	p-value	Z-value	p-value	Z-value	p-value
<i>Comet assay data</i>						
Low	-13.588	<0.0001	-30.993	<0.0001	-7.3321	<0.0001
Medium	-7.8555	<0.0001	-33.152	<0.0001	-7.5248	<0.0001
High	-9.6268	<0.0001	-35.460	<0.0001	-7.0866	<0.0001
VC	-11.332	<0.0001	-16.646	<0.0001	-7.3331	<0.0001
<i>CHS study – without censoring</i>						
Gender	9.0524	<0.0001	5.8756	<0.0001	2.1805	0.0146
BMI <sub>2</sub>	-4.6542	<0.0001	-9.6232	<0.0001	-0.7886	0.2152
BMI <sub>3</sub>	-3.2523	0.0006	-8.1637	<0.0001	-1.1624	0.1225
Hypertension	5.5865	<0.0001	5.2881	<0.0001	1.7644	0.0388
<i>CHS study – with censoring</i>						
Gender	0.1376	0.4453	4.8239	<0.0001	1.3061	0.0958
BMI <sub>2</sub>	-3.2394	0.0006	-0.4348	0.3319	-0.7489	0.2269
BMI <sub>3</sub>	-2.9269	0.0017	-2.0488	0.0202	-1.1571	0.1236
Hypertension	3.3776	0.0004	4.6403	<0.0001	2.0540	0.0110

**Table 4:** Combined model. Result of simulation study.

Full likelihood				Pairwise likelihood				Bayes				
Par	$\xi_1$	$\lambda$	$\alpha$	$\sigma$	$\xi_1$	$\lambda$	$\alpha$	$\sigma$	$\xi_1$	$\lambda$	$\alpha$	$\sigma$
True value												
# subjects: 20; $n_i$ : 10												
Est.	2.0928	0.2001	2.2774	0.2802	2.0997	0.1984	2.3226	0.2932	2.0910	0.2021	2.3814	0.2800
s.e.	1.0136	0.0284	0.6285	0.2171	0.9947	0.0278	0.5561	0.2132	1.0210	0.0294	0.7648	0.1259
s.e.(MC)	1.0102	0.0269	0.8491	0.1202	1.0054	0.0263	0.7845	0.1281	1.0025	0.0275	0.9594	0.0909
Bias	0.0928	0.0001	0.2774	0.2198	0.0997	0.0016	0.3226	0.2068	0.0910	0.0021	0.3814	0.2199
Rel.bias	0.0464	0.0005	0.1387	0.4396	0.0499	0.0077	0.1613	0.4136	0.0455	0.0106	0.1907	0.4399
# subjects: 40; $n_i$ : 10												
Est.	1.9949	0.1997	2.1142	0.2610	2.0107	0.1995	2.1337	0.2708	2.0029	0.2006	2.1448	0.2566
s.e.	0.7174	0.0196	0.3582	0.1485	0.7134	0.0196	0.3481	0.1709	0.7198	0.0199	0.3760	0.0963
s.e.(MC)	0.7049	0.0190	0.3899	0.1009	0.6952	0.0190	0.3956	0.1125	0.7106	0.0191	0.4304	0.0780
Bias	0.0051	0.0003	0.1142	0.2390	0.0107	0.0005	0.1337	0.2292	0.0029	0.0006	0.1448	0.2434
Rel.bias	0.0026	0.0016	0.0571	0.4780	0.0054	0.0026	0.0669	0.4583	0.0015	0.0031	0.0724	0.4869



**Table 5:** Multilevel combined model. Result of simulation study.

Full likelihood						Pairwise likelihood						Bayes				
Par	$\xi_1$	$\lambda$	$\alpha$	$\sigma_1^2$	$\sigma_2^2$	$\xi_1$	$\lambda$	$\alpha$	$\sigma_1^2$	$\sigma_2^2$	$\xi_1$	$\lambda$	$\alpha$	$\sigma_1^2$	$\sigma_2^2$	
	2	0.2	2	0.25	0.04	2	0.2	2	0.25	0.04	2	0.2	2	0.25	0.04	
# subj.(1):20; # subj.(2): 20; $n_i$ : 10																
Est.	2.0176	0.2144	1.8171	0.1275	0.0656	1.9493	0.2127	2.0326	0.4269	0.3814	1.9986	0.2029	2.0291	0.0728	0.0165	
s.e.	0.2302	0.0136	0.0824	0.0237	0.0241	0.2363	0.0138	0.1009	0.0476	0.0426	0.2244	0.0134	0.0989	0.0305	0.0095	
s.e.(MC)	0.2513	0.0157	0.0973	0.0554	0.0250	0.2501	0.0154	0.1091	0.1487	0.1392	0.2310	0.0148	0.0964	0.0253	0.0061	
Bias	0.0176	0.0144	0.1829	0.1225	0.0256	0.0507	0.0127	0.0326	0.1769	0.3414	0.0013	0.0029	0.0291	0.1771	0.0235	
Rel.bias	0.0088	0.0722	0.0914	0.4900	0.6404	0.0254	0.0634	0.0163	0.7077	8.5353	0.0007	0.0147	0.0146	0.7086	0.5865	
# subj.(1):20; # subj.(2): 40; $n_i$ : 10																
Est.	2.0047	0.2092	1.8047	0.1293	0.0647	2.0272	0.2066	1.9807	0.3768	0.3397	2.0094	0.2012	2.0192	0.0734	0.0129	
s.e.	0.1628	0.0124	0.0574	0.0168	0.0221	0.1611	0.0123	0.0720	0.0356	0.0282	0.1584	0.0125	0.0694	0.0286	0.0065	
s.e.(MC)	0.1734	0.0132	0.0952	0.0691	0.0250	0.1958	0.0143	0.0764	0.1184	0.1232	0.1584	0.0132	0.0723	0.0244	0.0042	
Bias	0.0047	0.0092	0.1953	0.1207	0.0247	0.0272	0.0066	0.0193	0.1268	0.2997	0.0094	0.0012	0.0192	0.1766	0.0271	
Rel.bias	0.0024	0.0463	0.0977	0.4829	0.6174	0.0136	0.0329	0.0096	0.5073	7.4930	0.0047	0.0059	0.0096	0.7063	0.6769	