Maastricht University | universiteit hasselt

2013 | School voor Informatietechnologie

DOCTORAATSPROEFSCHRIFT

Flexible modeling tools for hierarchical and incomplete data, with applications in comet assays and clinical trials

Proefschrift voorgelegd tot het behalen van de graad van doctor in de statistiek, te verdedigen door

Aklilu Habteab Ghebretinsae

Promoter: prof. dr. Christel Faes
Copromoter: prof. dr. Geert Molenberghs

D/2013/2451/50

Jury Members

Prof. dr. Christel Faes	-	Promotor: I-BioStat, University of Hasselt, Belgium.
Prof. dr. Geert Molenberghs	-	Copromotor: I-BioStat, University of Hasselt & KULeuven, Belgium.
Prof. dr. Marc Aerts	-	I-BioStat, University of Hasselt, Belgium.
Prof. dr. Niel Hens	-	I-BioStat, University of Hasselt & University of Antwerpen, Belgium.
Prof. dr. Geert Verbeke	-	I-BioStat, KULeuven & University of Hasselt, Belgium.
Prof. dr. Helena Geys	-	Janssen Pharmaceutica & University of Hasselt, Belgium.
Prof. dr. Dimitris Rizopoulos	-	Erasmus University, Rotterdam, the Netherlands.

Acknowledgements

First and foremost, I would like to thank the Almighty God who has guided me through the good and bad times of my life. without His grace and mercy, I wouldn't even be alive today let alone reach this level.

This work wouldn't have been possible without the support of many people. I am deeply grateful to all of those who supported me throughout my studies and my dissertation though I cannot name all of them. First, I owe my supervisors Prof. dr. Christel Faes and Prof. dr. Geert Molenberghs a gratitude for the patient guidance and advice they have provided me throughout my studies. I was fortunate to work under their supervision and to learn from them. I will be forever indebted for their never failing support. I am also grateful to the jury members for their feedback and suggestions on an earlier version of this dissertation. I would like to thank all of the people who collaborated with me over the years, including Prof. dr. Helena Geys, who provided me with data and arranged visits to Janssen Pharmaceutica, and Prof. dr. Niel Hens who worked with me on a research project. I am grateful to the secretaries of Center for Statistics, especially Ms. Martine Machiels, for their continued assistance.

As anyone can imagine graduate studies can be very demanding, and the Eritrean Christian community in Antwerpen has provided me needed escape and spiritual nourishment. They made me feel at home while away from home.

Throughout my education and my life, my family members, especially my deceased parents who showed me the right path and encouraged me to pursue my studies, have been supportive. I am forever grateful for their material and moral support. I am also thankful to all my friends for their encouragement and friendship.

Aklilu Habteab Ghebretinsae
Diepenbeek, 16 September 2013

Contents

List of Publications

This dissertation has been based on the following scientific papers:

Ghebretinsae, A. H., Faes, C. , Molenberghs, G., Geys, H. and Van der Leede, B. (2012). Joint modeling of hierarchically clustered and overdispersed non-gaussian continuous outcomes for comet assay data. *Pharmaceutical Statistics*, **11**,6, Page 449–455.

Ghebretinsae, A. H., Faes, C. , Molenberghs, G., De Boeck, M. and Geys, H.(2013). A Bayesian, Generalized Frailty Model for Comet Assays. *Journal of Biopharmaceutical Statistics*, **23**,3, Page 618–636.

Ghebretinsae, A. H., Molenberghs, G., Dmitrienko, A., Offen, W. and Sethuraman, G. (2013). Assessment of Type I Error Rate Associated with Dose Group Switching in longitudinal Alzheimer Trial. (Submitted to Journal of Biopharmaceutical Statistics, accepted).

De Preter, V., **Ghebretinsae, A. H.**, Abrahantes, J. C., Windey, K., Rutgeerts, P. and Verbeke, K. (2011). Impact of the Synbiotic combination Lactobacillus casei shirota and Oligofructose enriched Inulin on the Fecal Volatile Metabolite Profile in healthy subjects. *Molecular Nutrition and Food Reseach*, **55**, 5, Page 714–722.

Ghebretinsae, A. H., Faes, C. and Molenberghs, G. (2013). Gaussian variational approximation for GLMM with Overdispersion. (To be Submitted).

Ghebretinsae, A. H., Faes, C., Molenberghs, G. and Geys, H. (2013). Mixture Models for Comet assay. (To be Submitted).

Hens, N., **Ghebretinsae, A. H.**, Hardt, K., Van Damme, P., Van Herck, K. (2013). Model based estimates of long-term persistence of inactivated hepatitis A vaccine-induced antibodies in adults.(Submitted to Vaccine)

Overview of the Thesis

In this thesis, some modeling issues and design aspects that arise in toxicological studies and clinical trials are addressed. The text is structured in two parts. The first part of the thesis is motivated by a comet assay, a toxicological study design to assess DNA damage, which has been a standard tool in the pharmaceutical industry for the assessment of the safety of potential new drugs. In this part, a flexible modeling tool is proposed addressing the different modeling issues and an estimation technique is explored. In particular, various models, i.e., a flexible model for hierarchically clustered and overdispersed outcomes, the multivariate extension through a joint modeling technique, mixture models, zero-inflated models, and the use of Gaussian variational approximation are discussed in this part.

The second part is related to incomplete data in clinical trials. It is very common practice for patients to drop out from a study due to different reasons. Such missing data, in general, have a potential to affect/distort inferences drawn. Some trials allow for data-driven adaptation when the dropout rate is high. The second part of the thesis focuses on the impact of such data driven adaptations in some characteristics. In particular, the type I error rate associated with dose group switching is assessed when the primary analysis is in terms of a longitudinal outcome. The error rate is assessed through a simulation study, inspired by a clinical trial in Alzheimer's disease.

Part I

Flexible Modeling Tools for Hierarchically Clustered and Overdispersed Data with Applications in Comet Assays

Chapter 1

Introduction

In clinical trials and toxicological studies, measurements are taken to assess the efficacy and/or safety of a potential drug. Measurements are not always recorded on a continuous scale, also binary, count, and time to event outcomes are encountered. The standard approach of modeling such non-normal data is through the generalized linear model (GLM). They are the most common class of regression models used to analyze various types of variables (Nelder and Wedderburn 1972, McCullagh and Nelder 1989, Agresti 2002). The exponential family distributions provide elegant specifications of the models. The most well-known examples include linear regression, logistic regression, and Poisson regression. An important extension of these models is the generalized linear mixed model, by the inclusion of normally distributed random effects, allowing to account for multi-level structure in the data (Molenberghs and Verbeke 2005). A common issue with non-Gaussian data is overdispersion in the sense that the variability in the data is not well described by the distributional mean-variance relationship (Hinde and Demétrio 1998). This can happen both in the univariate and in the multi-level setting. One approach to account for overdispersion in a univariate generalized linear mixed model is by the use of a conjugate random effect, such as, for example the negative binomial (Breslow 1984, Lawless 1978) and beta-binomial model (Skellam 1948, Kleinman 1973). In a more recent publication, Molenberghs, Verbeke, and Demétrio (2007) proposed a similar approach to account for overdisperion in a multilevel setting, by the use of two random effects, a normally distributed random effect to accommodate for the hierarchy and some overdispersion, and a conjugate random effect to account for the overdispersion in the data. They introduced a new general modeling framework for the analysis of overdispersed multilevel data, which is

often referred to as the *combined model* (Molenberghs, Verbeke, and Demétrio 2007, Molenberghs *et al* 2010).

This thesis is partly motivated by a toxicological study design that is regularly encountered in pre-clinical research: the so-called comet assay (Ejchart and Sadlej-Sosnowska 2003, Lovell and Omori 2008). It is a sensitive method to assess DNA damage. During the last decade the assay gained widespread use in various areas and has emerged as a standard tool in the pharmaceutical industry for the assessment of the safety of potential new drugs. Typically, a comet assay is a single cell microgel electrophoresis method detecting DNA damage in any target tissue or organ of which a single cell suspension can be prepared. Visualization of this DNA migration (typical comet-like structures) is performed by a fluorescent dye. An image analysis system coupled to a microscope permits quantification of DNA damage at the single cell level. Three measures are commonly used: the tail migration (i.e., tail length), percentage tail intensity, and tail moment. Moreover, it exhibits higher-order hierarchies. In essence, the comet assay represents a hierarchical design with animals nested within doses, a number of slides per animals and several cells measured per slide. Comet measures from an animal are clearly not normally distributed but are rather asymmetric, skewed, bi- or multimodal, a mixture of different distributions, etc. While such data consist of non-Gaussian outcomes in a multi-level hierarchical structure, traditional analyses typically completely or partly ignore this hierarchical nature by summarizing measurements within a cluster. The comet assay study is presented in detail in Section 2.1.

In this thesis, different modeling issues that exist in the data are addressed: the nature of the outcome variables, the higher-order hierarchical structure, the overdispersion effect, the presence of excess zeros, as well as the multivariate structure of the data. A flexible modeling framework is thus needed to accommodate the above mentioned modeling issues. Moreover, an alternative estimation technique is proposed for such models.

In Chapter 2, different datasets that are used throughout this thesis are introduced. The first datasets correspond to the comet assay design and two sets of data resulting from a toxicology study on 1,2-Dimethylhydrazine dihydrochloride compound are presented. The second dataset is from a clinical trial in epilepsy, which has a count type of outcome. A third dataset from a toxicology study on Ethylene glycol with a binary time of outcome, is described as well.

In Chapter 3, the overview of models for non-Gaussian data, and the existing estimation techniques are presented. The probability model has to reflect the na-

ture of the data. In this chapter the different models for non-Gaussian are presented which range from the basic GLM, generalized linear mixed model (GLMM), different overdispersion models, to the recently proposed model for modeling for overdispersed and hierarchically clustered data. The various estimation techniques and the approximation techniques are also presented.

In Chapter 4, a flexible modeling approach for hierarchically clustered and overdispersed non-Gaussian outcome for comet assays is proposed, based on the combined model (Molenberghs *et al* 2010). Whereas a conjugate gamma random effect is used to account for the overdispersion of the data, both gamma and normal random effects are considered to account for the hierarchical structure of the data. In this chapter, the outcomes are modelled univariately. However, the comet assay data exhibit a multivariate structure.

In general, multivariate longitudinal or clustered data are commonly encountered in clinical trials and toxicological studies. Typically, there is no single standard endpoint to assess the toxicity or efficacy of the compound of interest, but co-primary endpoints are available to assess the toxic effects or the working of the compound. Modeling the responses jointly is thus appealing to draw overall inferences using all responses and to capture the association among the responses. In Chapter 5, a further extension to a multivariate setting with hierarchically clustered and overdispersed non-Gaussian outcomes is proposed for analysis of the comet assay data. The two outcomes are jointly analyzed by assuming that the normal random effects for both endpoints are correlated. The association structure between the response is analytically derived.

The overdispersion is accounted for through continuous conjugate random effects in Chapters 4 and 5. While it is convenient because of the conjugacy, misspecification of this distribution is possible. In Chapter 6, the use of mixture models is explored for comet data. A finite mixture of models as an alternative way to account for overdispersion, which is useful if the overdispersion is driven by subpopulations, and zero-inflated models, which are also mixture models, are considered. In addition, the use of a mixture of the conjugate distributions is also considered, where deviation from a single conjugate distribution is allowed for, while the property of conjugacy can still be employed to ease computations.

Another important aspect is the estimation technique for the combined model. The main difficulty with this kind of models is the computational complexity due to the intractable multivariate integrals, as is the case for GLMM that involve such integrals with no analytic expression. Different estimation methods for these models were

already proposed: estimation using partial marginalization (Molenberghs, Verbeke, and Demétrio 2007, Molenberghs *et al* 2010), estimation in the Bayesian framework (Ghebretinsae *et al* 2013), and an approximate estimation based on pseudo-likelihood (Effendi, Molenberghs, and Verbeke 2010). In Chapter 7, we will investigate the use of Gaussian variational approximation methods as a computationally fast estimation method for the combined model. A range of overdispersed non-gaussian mixed models will be investigated.

Chapter 2

Motivating Examples

This chapter introduces the datasets that are used throughout this thesis. In Section 2.1, we present the comet assay datasets. In Section 2.2, we introduce data from a clinical trial in epilepsy. The data from a toxicological study on Ethylene glycol is then introduced in Section 2.3.

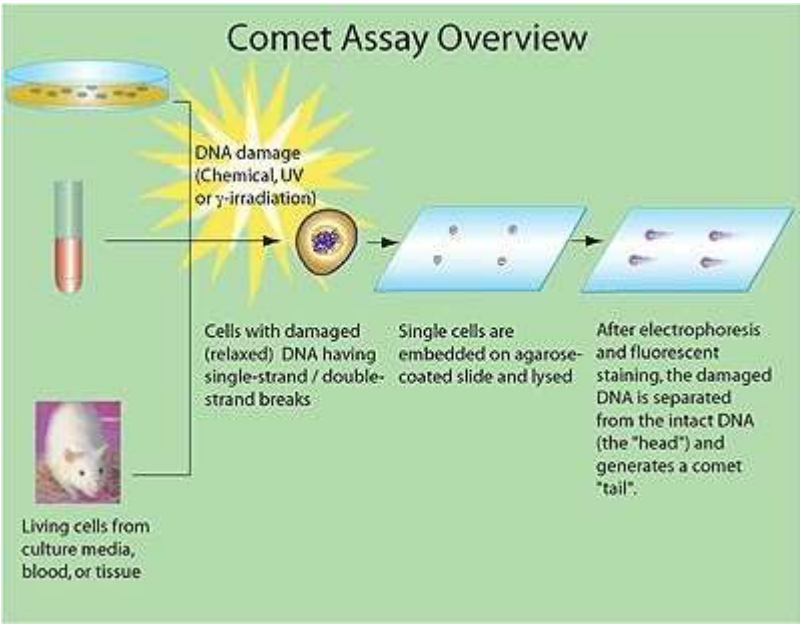
2.1 Comet Assay Data

A comet assay, regularly encountered in pre-clinical research, is a sensitive method to assess DNA damage. It was first developed by Ostling and Johanson in 1984 and later modified by Singh *et al* in 1988. It has since increased in popularity as a standard technique for evaluation of DNA damage/repair, biomonitoring and genotoxicity testing. During the last decade, the assay has emerged as a standard tool in the pharmaceutical industry for the assessment of the safety of potential new drugs. It involves the encapsulation of cells in a low-melting-point agarose suspension, lysis of the cells in neutral or alkaline conditions, and electrophoresis of the suspended lysed cells. Typically, it is a single cell microgel electrophoresis method detecting DNA damage in any target tissue or organ of which a single cell suspension can be prepared. Individual cells are embedded in a thin agarose gel on a microscope slide. All cellular proteins are then removed from the cells by lysing. The DNA is allowed to unwind under alkaline (pH>13.0) or neutral conditions. Following the unwinding, the DNA undergoes electrophoresis, allowing the broken DNA fragments or damaged DNA to migrate away from the nucleus. The resulting image obtained resembles a "comet" with a distinct head and tail. The head is composed of intact DNA, while

the tail consists of damaged (single-strand or double-strand breaks) or broken pieces of DNA. The extent of DNA liberated from the head of the comet is directly proportional to the amount of DNA damage. Visualization of this DNA migration is performed by a fluorescent dye. An image analysis system coupled to a microscope permits quantification of DNA damage at the single cell level. Three measures are commonly used: the tail migration (i.e., tail length), percentage tail intensity, and tail moment. Tail length is the distance from the perimeter of the comet head to the last visible point in the tail, percentage tail intensity is the percentage of DNA fragments present in the tail, while tail moment is the product of the amount of DNA in the tail and the mean distance of migration in the tail. In many experiments, the cells from a single animal are placed on a number of slides. Although there is no consensus among the experts as to the most appropriate statistical method and design (the number of slides and the replicates). Some studies (Wiklund and Agurell 2003, Smith *et al* 2008) indicate 3 slides and about 50 replicates/cells would be appropriate. A summary of the comet assay is presented in Figure 2.1.

The statistical analysis of such a comet assay is complicated because of several issues in the data. The comet assay represents a hierarchical design (Figure 2.2) with animals nested within doses, a number of slides per animals and several cells measured per slide. Comet measures from an animal are oftentimes not normally distributed. The complications that arise from the various non-normal distributions of comet endpoints are avoided in most standard analyses through the use of the central limit theorem. While the original data at the cell level may not be normally distributed, mean (or median) summaries at slide or animal level will be approximately normally distributed (given the typically large sample sizes) and are thus amenable to standard statistical analyses. Hierarchical or multilevel models make use of information on the various levels of variability but may be quite complex in terms of the distribution between cells of the same animals and difficult to interpret and explain. Their advantage, however, is that they provide estimates of the variability at each level and make use of the information at the cell level thus increasing the power of the study especially if the between-animal variability is not too large. Variability is expected between slides because of the variability in the handling of the different slides, and also variability between animals is expected, because of the individual-specific differences. Analyses on the same data indicate the importance of slide variability in contrast to the smaller rat variability (Ghebretinsae *et al* 2013).

The datasets resulted from a comet assay designed to assess the genotoxic potential of 1,2-Dimethylhydrazine dihydrochloride at different dose levels. Two datasets are



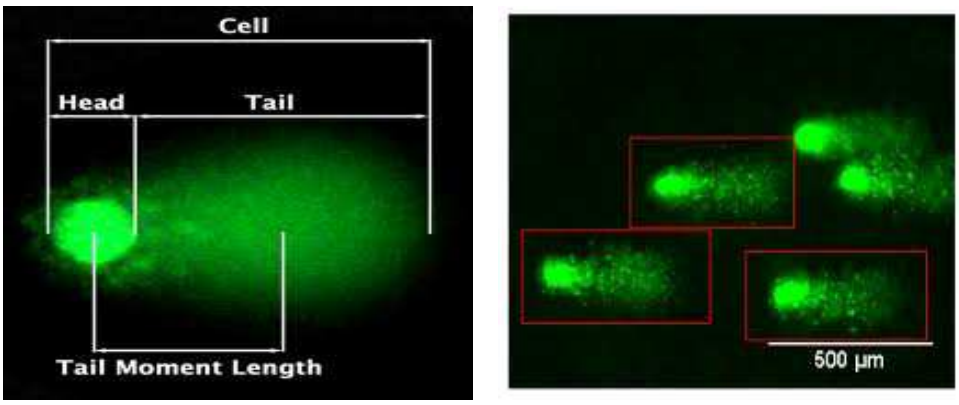


Figure 2.1: *Comet assay. Upper: overview of comet assay design. Lower: visualization of comets corresponding to the DNA damage of the cells.(Life Science: Comet Assay)*

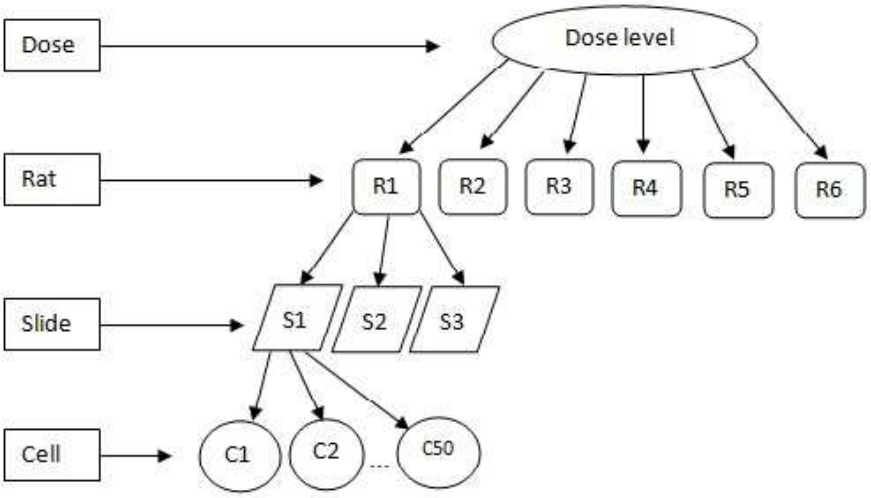


Figure 2.2: *Comet assay. Hierarchical structure of the design.*

considered. The first set has been studied in Ghebretinsae *et al* (2013). It refers to four groups of six male rats that received a daily oral dose of a compound in three dose levels (low, medium, and 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 3 hours after the last dose administration, their liver was removed, and processed for the comet assay. For each animal, a cell suspension is prepared. From each cell suspension, three replicate samples were prepared for scoring. Fifty randomly selected, non-overlapping cells per sample were then scored for DNA damage using a semi-automated scoring system. Thus, a total of 150 liver cells were scored per animal, on three slides. Generally, the toxicity level increases with the dose level. A summary of the data for tail intensity and tail length are represented in Figure 2.3. It indicate dispersion is more pronounced for tail length, due in part to the occurrence of zeros. We observe some extreme values at all dose levels.

The second set refers to four groups of six male rats that received a daily oral dose of a compound in three dose levels (low, medium, and high) or vehicle control. Also in the second datasets, a total of 150 liver cells were scored per animal. But here, a pronounced zero observations were observed in all dose groups. A summary of the data for tail length is represented in Figure 2.4.

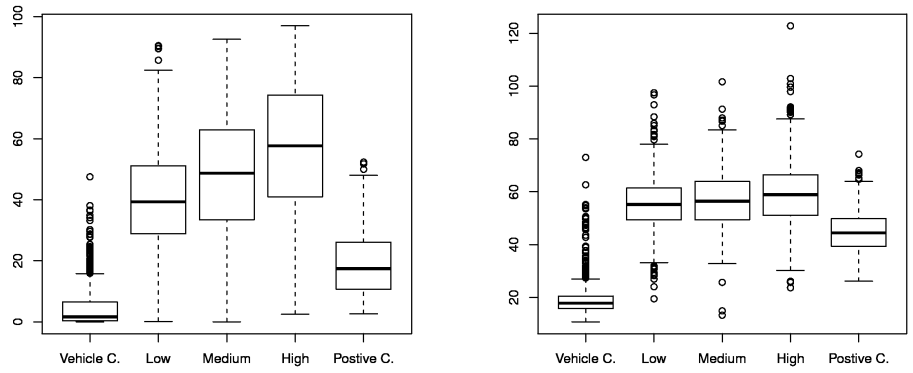


Figure 2.3: *Comet assay data 1. Box plots of tail intensity (left) and tail length (right) at each dose level*

2.2 A Clinical Trial in Epilepsy

The data considered here is obtained from a randomized, double-blind, parallel group multi-center study for the comparison of placebo with anti-epileptic drug (AED), in combination with one or two other AED's. The study is described in full detail in Faught *et al* (1996) and it is used in Molenberghs, Verbeke, and Demétrio (2007). The randomization of epileptic patients took place after a 12-week baseline period that served as a stabilization period for the use of AED's, and during which the number of seizures were counted. After that period, 45 patients were assigned to the placebo group, 44 to the active (new) treatment group. Patients were then measured on a weekly basis during 16 weeks, after which they were entered into a long term open extension study. Some patients were followed for up to 27 weeks. The outcome of interest is the number of epileptic seizures experienced during the last week, i.e., since the last time the outcome was measured. The key research question is whether or not the additional new treatment reduces the number of epileptic seizures. As a summary of the data, the distribution of the response is presented in the left panel of Figure 2.5 and the average profile for the two treatment groups over time is presented in the right panel of Figure 2.5. It produces a skewed distribution with largest observed value equal to 73 seizures at one particular week.

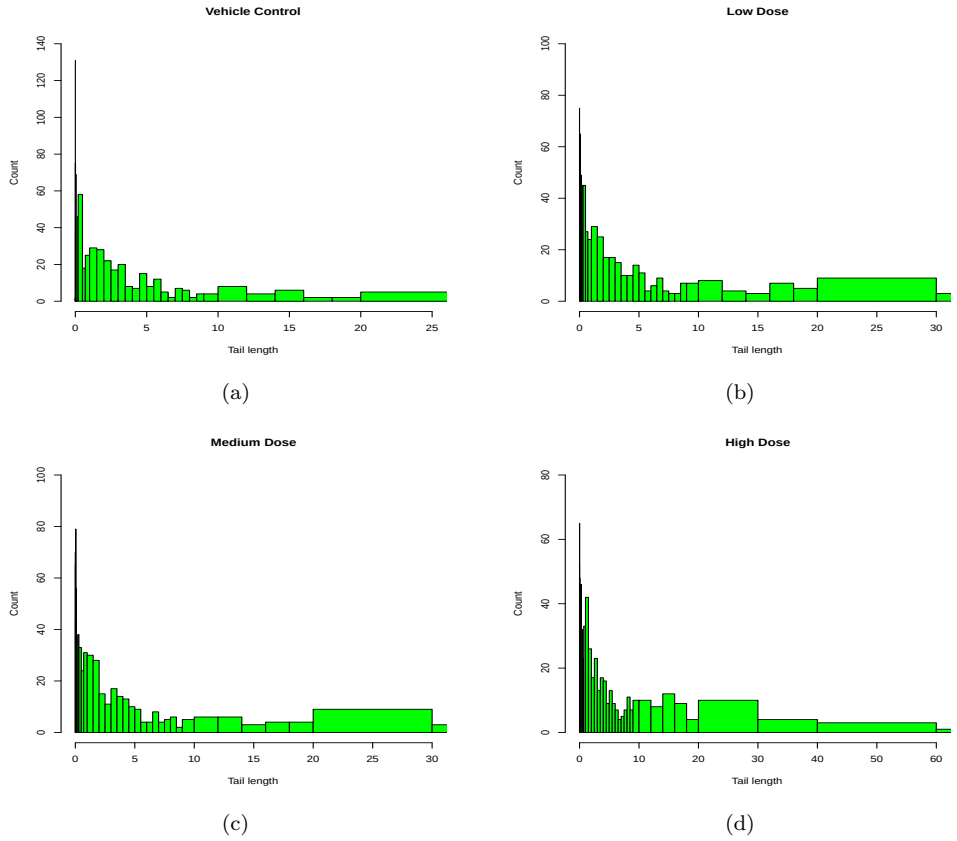


Figure 2.4: *Comet assay data 2. Distribution of tail length for each dose level*

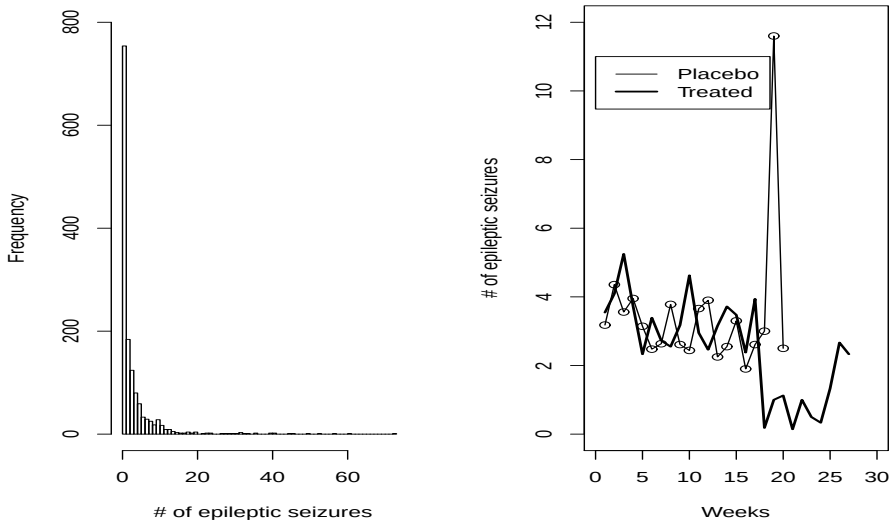


Figure 2.5: *Epilepsy data. The distribution of the response (left) and the average profile for the two treatment groups over time (right).*

2.3 Ethylene Glycol (EG) Data

The third dataset is from a toxicology study of ethylene glycol. Ethylene glycol (EG), also called 1,2-ethanediol is a high-volume industrial chemical with diverse applications. It is used to make antifreeze and de-icing solutions for cars, airplanes and boats, to make polyester compounds, and is used as a solvent in the paint and plastic industries. It is also used as an ingredient in photographic developing solutions, hydraulic brake fluids and in the formulation of several types of inks and many more. While EG may not be hazardous to humans in normal industrial handling, it can become dangerous when used at elevated temperatures or when ingested. Exposure to large amounts of ethylene glycol can damage the kidneys, heart, and nervous system. In addition, ingestion of antifreeze products, which consist for approximately 95 % of EG, is toxic and may result in death. The data resulted from a study in which timed-pregnant CD-1 mice were dosed by gavage with EG in distilled water as described by Price *et al* (1985). Dosing occurred during the period of major organogenesis and structural development of the foetuses (gestational days 6 through 15). The doses

were administered at 0, 750, 1500, or 3000 mg/kg/day, with 25, 24, 23, and 23 timed-pregnant mice randomly assigned to each of these dose groups, respectively. The interest here is to assess the toxicity of this chemical at the different dose levels based on a binary outcome, whether the foetus is malformed or not. Summary of the data is presented in Figure 2.6. We observe a general trend of increasing toxicity with dose level.

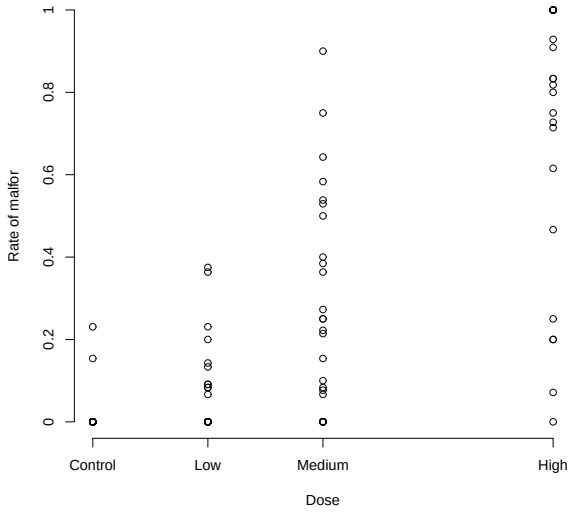


Figure 2.6: *EG data. Scatter plot of the rate of malformation in a litter, as a function of dose.*

Chapter 3

Methodological Background

Each of the datasets described in the previous chapter deals with correlated non-normal data. In addition, these are hierarchically structured with two (for Epilepsy and EG data) or three (for comet data) levels in the hierarchy. In building a flexible model for such a complex data setting, some choices have to be made in regard to the appropriate probability model and estimation method. The probability model needs to reflect the nature of the data. One of these is the type of outcome: count, binary, time to event, or continuous. Accordingly, an appropriate probability distribution of the data collected need be considered. The other aspect is the data structure: whether the data have hierarchical, clustered structure or not. As a result, different modeling approaches can be considered: random-effect, marginal, or conditional models. The choice between these may mainly depend on the research question one wants to answer. However, the computational ease and the availability of software tools also plays a role in the choice. In addition, when the data is highly skewed and/or an overdispersion issue exists, models that accommodate these issues need be considered.

Another issue is the estimation method. Based on the model formulation, an appropriate estimation method has to be chosen. Estimation methods range from full likelihood to approximating methods such as pseudo-likelihood and quasi-likelihood. Likelihood methods enjoy many desirable properties, such as efficiency under appropriate regularity conditions and the ability to calculate functions of interest based on the proposed parametric model (Edwards 1971). However, the estimation of the parameters can be computationally intensive. As a result, alternative estimation techniques are often of interest. But also, estimation can be done in a Bayesian framework,

as an alternative to the likelihood framework.

In this chapter, a review of the existing models for the analysis of such non-normal data as well as the estimation techniques is given. Section 3.1 presents an the overview of models for non-normal data with overdispersion. The various techniques are reviewed in Section 3.2.

3.1 Overview of Models for Modeling Non-normal Data

3.1.1 Standard Generalized Linear Models

A standard approach for modeling non-normal data is the generalized linear model (GLM). Let us assume that a random variable Y follows an exponential family distribution if the density is of the form:

$$f(y) = f(y|\eta, \phi) = \exp\left\{\phi^{-1}[y\eta - \psi(\eta)] + c(y, \phi)\right\}, \tag{3.1}$$

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

The first two moments follow from the function $\psi(\cdot)$ (Molenberghs and Verbeke 2005) and are given by:

$$E(Y) = \mu = \psi'(\eta), \tag{3.2}$$

$$\text{Var}(Y) = \sigma^2 = \phi\psi''(\eta), \tag{3.3}$$

An important implication is that, in general, the mean and variance are related through $\sigma^2 = \phi\psi''[\psi^{-1}(\mu)] = \phi v(\mu)$, with $v(\cdot)$ the so-called variance function, describing the mean-variance relationship. This relationship exists in the exponential family for binary, count, and time-to-event data. The normal model is a special one, in particular because the overdispersion parameter is needed to allow for a variance other than unity. As a result, the mean-variance relationship, is absent for this model, but present for all others.

In a regression context, where one wishes to explain variability between outcome values based on measured covariate values, the model needs to incorporate covariates. This leads to so-called generalized linear models. Let $Y_1, \ldots, Y_N$ be a set of

independent outcomes, and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ represent the corresponding p-dimensional vectors of covariate values. It is assumed that all Y_i have densities $f(y_i|\eta_i, \phi)$, which belong to the exponential family, but a different natural parameter η_i is allowed per observation. Specification of the generalized linear model is completed by modeling the means μ_i as functions of the covariate values. More specifically, it is assumed that $\mu_i = h(\eta_i) = h(\boldsymbol{x}_i'\boldsymbol{\xi})$, for a known function $h(\cdot)$, and with $\boldsymbol{\xi}$ a vector of p fixed, unknown regression coefficients. Usually, $h^{-1}(\cdot)$ is called the link function. In most applications, the so-called natural link function is used, i.e., $h(\cdot) = \psi'(\cdot)$, which is equivalent to assuming $\eta_i = \boldsymbol{x}_i'\boldsymbol{\xi}$. Hence, it is assumed that the natural parameter satisfies a linear regression model.

3.1.2 Overdispersion Models

As presented in Molenberghs *et al* (2010), the standard Bernoulli, Poisson, and exponential models force the mean and variance functions to depend on a single parameter. However, comparing the sample average with the sample variance might already reveal in certain applications that this assumption is not in line with a particular set of data, for count and time-to-event data, for example. Therefore, a number of extensions have been proposed. Hinde and Demétrio (1998ab) provide general treatments of overdispersion. The Poisson case received particular attention by Breslow (1984) and Lawless (1987). Molenberghs and Verbeke (2005) mentions various model-based approaches that accommodate overdispersion, including the betabinomial model (Skellam 1948), the Bahadur model (1961), the multivariate probit model (Dale 1986, Molenberghs and Lesaffre 1994), and certain versions of the generalized linear mixed model (Breslow and Clayton 1993).

A straightforward and commonly encountered step is to allow the overdispersion parameter $\phi \neq 1$, so that $\text{Var}(Y) = \phi v(\mu)$. This is in line with the moment-based approach mentioned in the previous section, but can also be engendered by fully parametric assumptions. Another way forward is through a random effect. Assuming no particular distributional form for the random effects gives rise to a semi-parametric specification. Assuming full distributional assumptions about the random effects may be advantageous though. In that case, the common choices are the beta distribution for binomial outcomes and the gamma distribution for Poisson and Weibull outcomes. Generally, the model is made up of two components: a distribution for the outcome, given a random effect $f(y_i|\theta_i)$, and a distribution for the random effect, $f(\theta_i)$. If the

two are combined, it produces the marginal model:

$$f_i(y_i) = \int f(y_i|\theta_i) f(\theta_i) d\theta_i. \tag{3.4}$$

Considering gamma random effects for the Poisson model, beta random effects for binomial data, and gamma random effects for the Weibull model may look a disparate collection. However, they are bound together by the property of conjugacy, in the sense of Cox and Hinkley (1974, p. 370) and Lee, Nelder, and Pawitan (2006, p. 178). Informally, conjugacy refers to the fact that the hierarchical and random-effects densities have similar algebraic forms. Conjugate distributions produce a general and closed-form solution for the corresponding marginal distribution.

In the case of binary data, the marginal model is the familiar beta-binomial model. For count data, the negative-binomial model results. Unlike in the binary case, univariate counts are able to violate the mean-variance relationship of the Poisson distribution, hence the great popularity of this and other types of models for overdispersion.

The parameters α and β in the beta and gamma distributions are not always jointly identified. It is therefore customary to impose restrictions, such as setting one of them equal to a fixed value, e.g., $\alpha = 1$, or constraining their mean or variance, etc.

3.1.3 Generalized Linear Mixed Models

The generalized linear mixed model (GLMM; Breslow and Clayton 1993, Molenberghs and Verbeke 2005) is the most frequently used random-effects model in the context of non-Gaussian repeated measurements. It is a relatively straightforward extension of the generalized linear model for independent data (Section 3.1.1) to the context of hierarchically organized data on one hand and the linear mixed model (Verbeke and Molenberghs 2000) on the other hand. A wide range of software tools is available for fitting such models. Let Y_{ij} be the j^{th} outcome measured for cluster (subject) i, $i = 1, \ldots, N, j = 1, \ldots, n_i$, and stack the n_i measurements into a vector $\boldsymbol{Y_i}$. Assume that, in analogy with Section 3.1.1, conditionally upon q-dimensional random effects $\boldsymbol{b_i} \sim N(0, D)$, the outcomes Y_{ij} are independent with densities of the form:

$$f_i(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}, \phi) = \exp\left[\phi^{-1}[y_{ij}\eta_{ij} - \psi(\eta_{ij})] + c(y_{ij}, \phi)\right], \tag{3.5}$$

with

$$\psi'(\eta_{ij}) = \mu_{ij} = E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}) = h(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}), \tag{3.6}$$

for a known function $h(\cdot)$ and $h^{-1}(\cdot)$ is the link function, with $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ p-dimensional and q-dimensional vectors of known covariate values, with $\boldsymbol{\xi}$ a p-dimensional vector of unknown fixed regression coefficients, and with ϕ a scale (overdispersion) parameter. Finally, let $f(\boldsymbol{b_i}|D)$ be the density of the $N(0, D)$ distribution for the random effects $\boldsymbol{b_i}$.

3.1.4 Combined Model

Integrating both the overdispersion random effects as well as the normal random effects into the generalized linear model framework, Molenberghs *et al* (2010) proposed an exponential family model to accommodate simultaneously the clustering and overdispersion effects. It extends the generalized linear mixed model by the use of conjugate random effect for overdispersion. The general model family proposed for modeling overdispersed and correlated data is given by:

$$f_i(y_{ij}|\boldsymbol{b_i}, \theta_{ij}, \boldsymbol{\xi}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\}, \tag{3.7}$$

for outcome Y_{ij} on subject $i = 1, \ldots, N$ at occasion $j = 1, \ldots, n_i$. The unknown parameters λ_{ij} and ϕ are termed natural parameter and scale parameter, respectively. The term $c(y_{ij}, \phi)$ is the normalizing constant. The function $\psi(\cdot)$ is a known function with the property that $E[y_{ij}|\boldsymbol{b_i}, \theta_{ij}, \boldsymbol{\xi}] = \psi'(\lambda_{ij})$ and $\text{Var}(y_{ij}|\boldsymbol{b_i}, \theta_{ij}, \boldsymbol{\xi}) = \phi\psi''(\lambda_{ij})$. Model specification proceeds by assuming that the conditional mean of Y_{ij} is given by

$$E[Y_{ij}|\theta_{ij}, \boldsymbol{b_i}] = \mu_{ij}^c = \theta_{ij}\kappa_{ij}, \tag{3.8}$$

where $\theta_{ij} \sim \mathcal{G}_{ij}(\xi_{ij}, \sigma_{ij}^2)$ for some distribution $\mathcal{G}_{ij}$ with mean ξ_{ij} and variance σ_{ij}^2 and $\kappa_{ij} = g(\eta_{ij}) = g(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i})$ for some function g and $\boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$. The random variable θ_{ij} is used to account for the overdispersion in the data, while the random effect in κ_{ij} accounts for the clustered or hierarchical structure of the data. The two parameters η_{ij} and λ_{ij} refer to the linear predictor and/or the natural parameter. The basic difference is that λ_{ij} encompasses the random variables θ_{ij}, whereas η_{ij} refers to the 'GLMM part' only.

Most often, but not strictly necessary, it is assumed that the two sets of random effects, $\boldsymbol{\theta}_i$ and $\boldsymbol{b_i}$, are independent of each other. Regarding the components θ_{ij} of $\boldsymbol{\theta}_i$, three useful special cases result from assuming that: (1) they are independent; (2) they are correlated, implying that the collection of univariate distributions $\mathcal{G}_{ij}(\xi_{ij}, \sigma_{ij}^2)$ needs to be replaced with a multivariate one; and (3) they are equal to each other,

useful in applications with exchangeable outcomes Y_{ij} (see Molenberghs *et al* 2010 for further discussion).

Parameterization (3.8) is such that the random effects θ_{ij} capture overdispersion, and are formulated directly at the mean scale, whereas κ_{ij} can be considered the generalized linear mixed model component.

An important concept in regard to computational efficiency is *conjugacy*, in the sense of Cox and Hinkley (1974, p. 370) and Lee, Nelder, and Pawitan (2006, p. 178). Conjugacy refers to the fact that the hierarchical and random-effects densities have similar algebraic forms. Conjugate distributions produce a general and closed-form solution for the corresponding marginal distribution. Molenberghs *et al* (2010) adapted conjugacy to the situation where both normal and overdispersion random effects are included. For further explanation see Molenberghs *et al* (2010).

A set of three types of outcomes are further considered: time-to-event, count and binary. A Poisson model will be considered for counts, a Weibull-exponential model for time-to-event and a logistic model for binary outcomes.

3.1.4.1 Poisson-type Models for Count Data

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

$$\begin{aligned} Y_{ij} &\sim \text{Poisson}(\lambda_{ij}), && (3.9)\\ \lambda_{ij} &= \theta_{ij}\exp(\boldsymbol{x}_{ij}{}'\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}), && (3.10)\\ \boldsymbol{b_i} &\sim \text{Normal}(0, D), && (3.11)\\ \theta_{ij} &\sim \text{Gamma}(\alpha, \beta). && (3.12) \end{aligned}$$

It is a Poisson-gamma-normal model or, equivalently, a negative-binomial-normal model. It is implicitly assumed that the components θ_{ij} of $\boldsymbol{\theta}_i$ are independent. This is natural in many cases in the sense that the $\boldsymbol{b_i}$ will induce association between repeated measurements, with then the θ_{ij} taking care of additional dispersion. For a dependent θ_{ij}, then one could choose, for example, multivariate extensions of the gamma model (Gentle 2003).

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

3.1.4.2 Weibull- and Exponential-type Models for Time-to-event Data

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

$$f(y_{ij}|\theta_{ij}, \boldsymbol{b_i}) = \lambda\rho\theta_{ij}y_{ij}^{\rho-1}e^{\eta_{ij}}e^{-\lambda y_{ij}^{\rho}\theta_{ij}e^{\eta_{ij}}}, \tag{3.13}$$

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}, \tag{3.14}$$

$$f(\theta_{ij}) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\theta_{ij}^{\alpha-1}e^{-\theta_{ij}/\beta}, \tag{3.15}$$

$$f(\boldsymbol{b_i}) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}}e^{-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}}. \tag{3.16}$$

A few comments are in place. First, it is implicit that the gamma random effects are independent. This need not be the case and, like in the Poisson case, extension via multivariate gamma distributions is possible. Second, setting $\rho = 1$ leads to the special case of an exponential time-to-event distribution. Third, it is evident that the classical gamma frailty model (i.e., no normal random effects) and the Weibull-based GLMM (i.e., no gamma random effects) follow as special cases. Fourth, owing to the conjugacy and the following property of the gamma distribution:

$$\frac{1}{\kappa}f(\theta|\alpha,\beta) = \frac{1}{\kappa}\frac{1}{\beta^{\alpha}\Gamma(\alpha)}\theta^{\alpha-1}e^{-\theta/\beta}, \tag{3.17}$$

$$= \frac{1}{(\kappa\beta)^{\alpha}\Gamma(\alpha)}(\kappa\theta)^{\alpha-1}e^{-(\kappa\theta)/(\kappa\beta)}, \tag{3.18}$$

$$= f(\kappa\beta|\alpha,\kappa\beta). \tag{3.19}$$

strong conjugacy applies. This is typically considered for the exponential model, but it holds for the Weibull model too, merely by observing, that the Weibull model is nothing but an exponential model for the random variable Y_{ij}^{ρ}. It is equally possible to derive this result by merely re-writing the factor $\phi = \lambda\kappa$. Fifth, the above expressions are derived for a two-parameter gamma density. It is customary in a gamma frailty context (Duchateau and Janssen 2007) to set $\alpha\beta = 1$, for reasons of identifiability. In this case, (3.15) is replaced by

$$f(\theta_{ij}) = \frac{1}{\left(\frac{1}{\alpha}\right)^{\alpha}\Gamma(\alpha)}\theta_{ij}^{\alpha-1}e^{-\alpha\theta_{ij}}, \tag{3.20}$$

Alternatively, assuming $\alpha = 1$ and $\beta = 1/\delta$, one could write

$$f(\theta_{ij}) = \delta e^{-\delta\theta_{ij}}, \tag{3.21}$$

implying that the gamma density is reduced to an exponential one.

3.1.4.3 Bernoulli-type Models for Binary Data

Similar to the Poisson case, a natural binary-data counterpart is:

$$\begin{aligned} Y_{ij} &\sim \text{Bernoulli}(\pi_{ij} = \theta_{ij}\kappa_{ij}), && (3.22)\\ \kappa_{ij} &= \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i})}{1 + \exp(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i})}, && (3.23)\\ \boldsymbol{b_i} &\sim \text{Normal}(0, D), && (3.24)\\ \theta_{ij} &\sim \text{Beta}(\alpha, \beta). && (3.25) \end{aligned}$$

In comparison to the longitudinal Poisson case, the longitudinal binary case appears to defeat closed-form solutions and strong conjugacy. However, this hinges on the fact that we employ the logit link. In spite of it being a very natural choice in the univariate case, it does not combine very nicely with normal random effects. This is known already from the GLMM framework for binary data.

3.2 Estimation Techniques

Based on the model formulation, an appropriate estimation method should be chosen. Estimation methods range from full likelihood to pseudo-likelihood, and quasi-likelihood.

3.2.1 Maximum Likelihood Estimation

In general, for a fixed set of data and underlying statistical model, the method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the selected model with the observed data. Random effects models can be fitted by maximization of the marginal likelihood, obtained by integrating out the random effects. In the

generalized linear mixed model, the likelihood contribution of subject i becomes:

$$f_i(y_{ij}|\beta, D, \phi) = \int \prod_{j=1}^{n_i} f_i(y_{ij}|\boldsymbol{b_i}, D, \phi) f(\boldsymbol{b_i}) d\boldsymbol{b_i}, \tag{3.26}$$

from which the likelihood for β,D, and ϕ is derived as:

$$L(\beta, D, \phi) = \prod_{1=1}^{N} f_i(y_{ij}|\beta, D, \phi) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_i(y_{ij}|\boldsymbol{b_i}, D, \phi) f(\boldsymbol{b_i}) d\boldsymbol{b_i} \tag{3.27}$$

In a similar way, the combined model can be fitted by maximizing the marginal model, obtained by integrating out both random effects. Maximum likelihood methods enjoy many desirable properties, such as efficiency under appropriate regularity conditions. But it can be unattractive due to excessive computational requirements. For example, with multivariate exponential family models, it can have a cumbersome expression, rendering it hard to evaluate (Arnold and Strauss 1991). Several suggestions have been made to overcome this problem, such as Monte Carlo integration (Tanner 1991). For example, Geyer and Thompson (1992) use Markov Chain Monte Carlo simulations to construct Monte Carlo approximation to the analytically intractable likelihood. Arnold and Strauss (1991) proposed the use of a so called pseudo-likelihood (PL) function.

3.2.2 Approximate Estimation Techniques

The problem in maximizing (3.27) is the presence of the N-integrals over the q-dimensional random effects $\boldsymbol{b_i}$. In some special cases, these integrals can be worked out analytically. For example, this has been done for linear mixed models for continuous outcomes (Molenberghs 2005). In general, no analytic expressions are available for the integrals and numerical approximation is needed. There is a large statistical literature on various methods to do so. The numerical approximations can be subdivided in: (1) those that are based on the approximation of the integrand; (2) those on the approximation of the data and; (3) those that are based on the approximation of the integral itself.

3.2.2.1 Approximation of the Integrand

The first approach is through the approximation of the integrand. When the integrands are approximated, the goal is to obtain a tractable integral such that a closed

form expressions can be obtained, making the numerical maximization of the approximate likelihood feasible. Several methods have been proposed, but basically all come down to Laplace-type approximations of the function to be integrated (Molenberghs 2005). Recently, Ormerod and Wand (2012) introduced variational approximation in the statistical modeling framework. It approximates the integrand by introducing a set of approximating densities in such a way as to make their evaluation tractable. It will be discussed in Chapter 7.

3.2.2.2 Approximation of the Data

Another approach is based on a decomposition of the data into the mean and an appropriate error term, with a Taylor series expansion of the mean, which is a non-linear function of the linear predictor. All methods in this class differ in the order of the Taylor approximation and the point around which the approximation is expanded. Penalized Quasi- Likelihood (PQL) and Marginal Quasi- Likelihood (MQL) are examples of techniques based on approximation of the data. PQL is obtained by/from maximizing a quasi-likelihood function which only involves first- and second-order conditional moments, augmented with a penality term on the random effects. MQL is also very similar to PQL in the sense that it also depends on first- and second-order conditional moments, but now evaluated in the marginal linear predictor instead of the conditional linear predictor. The difference is MQL does not incorporate the random effect $\boldsymbol{b_i}$.

3.2.2.3 Approximation of the Integral

Where the above approximation methods fail, approximations to the integral, that is, numerical integration, is useful. Approxmation of the integral through Gaussian quadrature assumes that $\int f(z)\Phi(z)dz$ is approximated by:

$$\int f(z)\Phi(z)dz \approx \sum_{h=1}^{Q} w_h f(z_h).$$

The integral is approximated by a weighted sum evaluated at Q values z_h, called quadrature points. The quadrature points are the solutions to the Q^{th} order hermite polynomial. w_h are appropriately chosen weights. In the simple setting of univariate integration, the approximation consists of subdividing the integration region in intervals, and approximating the area under the curve by the sum of the areas of the

so-obtained rectangles. In general, the higher Q, the smaller the width of the intervals and the better the approximation. The quadrature points z_h are independent of the function $f(z)$, and as a result the z_h may or may not lie in the region of interest. In such a case, for a small value of Q, the quadrature points z_h can be inappropriate. In that case, it might be useful to rescale and shift the quadrature points such that more points lie in the region of interest. In the adaptive Gaussian quadrature the quadrature points are centered and scaled as if $f(z)\Phi(z)$ would follow a normal distribution.

3.2.3 Bayesian Estimation

When random-effects models are used, the likelihood function involves the integration over the random-effects distribution. In case responses are normally distributed, the marginal likelihood can be derived analytically. However, this property of normal models does not extend to the case of non-normal distributions, where in general, no closed forms are available. Estimation methods then either employ an approximation of the integrand or uses numerical integration techniques. The need for complex numerical integration can be avoided by casting the random-effects model into a Bayesian framework, and resorting to the Gibbs sampler (Zeger and Karim 1991).

In the Bayesian framework, the unknown parameters are estimated by the posterior mean. It is typically done by taking random draws from a posterior density using Markov chain Monte Carlo simulation (MCMC), particularly Gibbs sampling. The basic idea of Gibbs sampling is to partition the set of unknown parameters and then estimate them one at a time or in a group, conditional on all others. The Gibbs sampler starts with initial values for all parameters and then updates them in turn, giving each a random estimate based on the data and the current guess of the other parameters in the model (Gelman and Hill 2006). The sample averages are taken as the posterior means of the parameters of interest. A model selection procedure is needed in order to compare between models and to select the best fitting model. Goodness-of-fit and complexity of the models can be assessed using the deviance information criterion (DIC) as proposed by Spiegelhalter *et al* (1998, 2002) and recently used by Erkanli *et al* (2000), Rahmann *et al* (1999) and Gelfand *et al* (2000) for model selection within the Bayesian framework.

3.2.3.1 Deviance Information Criterion (DIC)

The deviance is defined as the posterior distribution of the log likelihood: $D = -2\log(p(y|\beta))+2\log f(y)$, where $p(y|\beta)$ is the posterior density, with $f(y)$ a standardizing term that does not affect model comparison. The goodness of fit of the model is then summarized by the posterior expectation of the deviance: $E_{\beta|y}[D]$. Spiegelhalter *et al* (1998, 2002) suggested to measure the complexity of the model by the difference between the posterior expectation of the deviance ($\bar{D}$) and the deviance evaluated at the posterior expectation of β ($\hat{D}$), that is:

$$pD = E_{\beta|y}[D] - D(E_{\beta|y}[\beta]), \quad (3.28)$$

$$= \bar{D} - \hat{D}, \quad (3.29)$$

where pD can be interpreted as the effective number of parameters in the model. These are combined to give the overall DIC:

$$DIC = \bar{D} + pD,$$

where the first term represents the goodness of fit and the second term represents the model complexity (the effective number of parameters). Smaller values of DIC indicate a better fitting model.

Chapter 4

Univariate Model for Hierarchically Clustered and Overdispersed Outcomes: Comet Assay Data

As explained in the previous chapters, the statistical analysis of a comet assay is complicated because of several issues in the data. In this chapter, a method is proposed accounting for different challenges: the multi-level structure of the data, the type of data, and the skewness of the outcome of interest.

In many protocols, the cells from a single animal are placed on a number of slides. Each cell is then investigated for DNA damage by measuring the tail length and tail intensity of the comet. Because variability is expected between slides and between animals, this needs to be taken into account in the statistical analysis. This results in three-level hierarchies, with clustering at the animal and slide level (Figure 2.2).

Moreover, exploration of the distribution of the gathered data and previous work in this area indicate that the distribution for the responses (tail length and tail intensity) are asymmetric (Lovell and Omori 2008). This is often completely or partly ignored in traditional analyses. The standard approach of modeling non-normal data, such as the tail intensity and tail length in the comet assay, is using a generalized linear model (e.g., a Weibull model). The generalized linear model framework (McCullagh

and Nelder 1989) is a very rich one. Nevertheless, as already discussed in Chapters 1 and 3, many standard members of the family may exhibit overdispersion due to a prescribed relationship between mean and variance. For example, in the exponential and Weibull cases, there is a quadratic relationship between them. This is why many proposals have been made to extend the models such that they can deal with so-called overdispersion, which is taken to mean that the actual relationship between mean and variance is different from the one prescribed.

Here, a random effects model is proposed accounting for both the overdispersion in the data and the hierarchical design of the assay. Random effects are broadly used to analyze outcomes collected in a repeated-measures, longitudinal, clustered, or multivariate fashion. But as mentioned in previous chapters, random effects can also be used to accommodate the overdispersion in the data. For example, when parameters in the Weibull model are thought of as being random and each observation is drawn from a different Weibull distribution, this would lead to an overdispersed Weibull model. An overview is given in Molenberghs *et al* (2010). Random effects are frequently assumed to be normal, but they can take various distributional forms, such as beta random effects with binomial data, gamma random effects with count data, etc. An illustrious counterexample is time-to-event data where gamma random-effects, usually termed gamma frailties, are in common use. Molenberghs *et al* (2010) proposed an extended framework where both types of random effects are considered simultaneously, so as to deal, at the same time, with overdispersion on the one hand and data hierarchies on the other.

Arguably, such model development, while requiring additional work, is necessary for a number of reasons. First, Molenberghs, Verbeke, and Demétrio (2007) showed that classical generalized linear mixed models (GLMM) can be inadequate to model, at the same time, overdispersion and data hierarchies. Precisely, they modeled repeatedly measured epileptic-seizure data and found that the more conventional GLMM exhibited inferior fit, but also that two types of inferences were incorrect under the simpler model: (1) the correlation between repeated measures was substantially overestimated with the GLMM and (2) the treatment effect with the GLMM was found significant whereas the extended model showed that there was no treatment effect at all. Thus, the spurious treatment effect was entirely a consequence of model misspecification. Second, the design considered here is even more complex, with various hierarchical levels; it is generally inappropriate to consider a model that does not fully accommodate the design. Third, even if the model could be simplified to a more conventional model, this cannot be uncovered without considering a more general

model. Thus, the model development proposed here can be used additionally as a goodness-of-fit tool for, say, the GLMM.

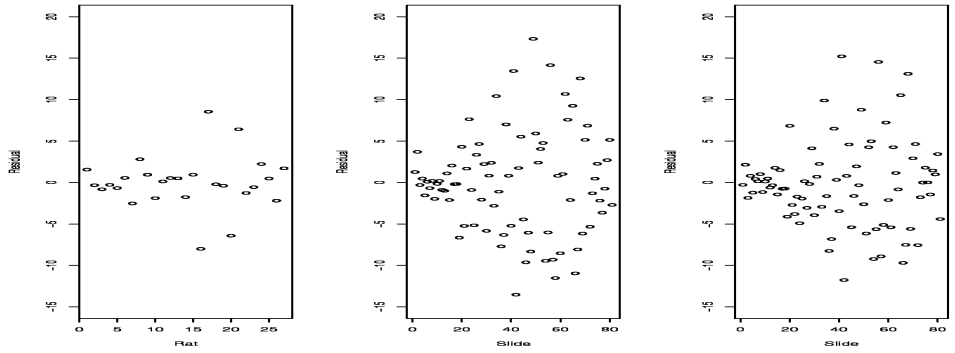
Here, we will focus on the specific case of a non-negative continuous outcome in view of the comet assay. Whereas Molenberghs *et al* (2010) considered a two-level hierarchy in the form of repeated measures on the same subject, the comet assay data exhibit higher-order hierarchies.

Also, these authors considered maximum likelihood estimation, but here we rather propose a Bayesian approach. Not only does it have computational advantages, it allows to take relevant information from preceding studies into account, a so-called Bayesian learning approach. The interest here is to see the toxicity of 1,2-Dimethylhydrazine dihydrochloride at the different dose levels (low, medium, and high) using the appropriate distribution and taking in to account the complete hierarchical nature. This work has been published in Ghebretinsae *et al* (2013).

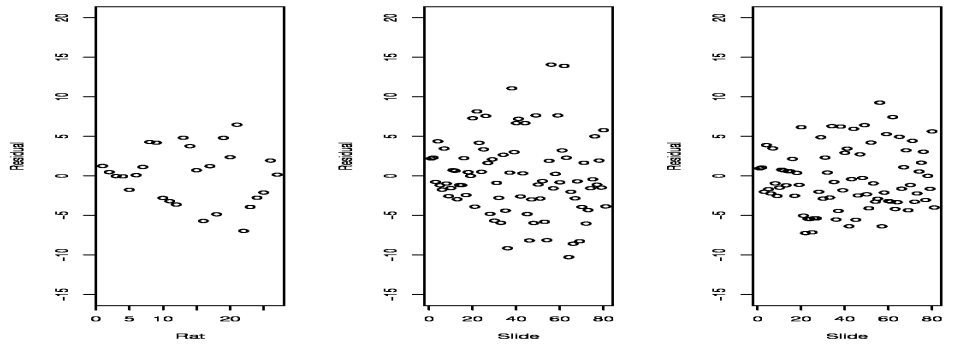
In this chapter, data are explored and the traditional analyses are presented in Section 4.1. Section 4.2 presents the framework for combined overdispersion and hierarchical random effects with non-Gaussian continuous outcome. Estimation methodology is discussed in Section 4.3. The effect of overdispersion and clustering is illustrated in Section 4.4. The data are analyzed in Section 4.5.

4.1 Data Exploration and Traditional Models

Although for the purpose of comparison across studies, statistical analyses are commonly performed on percentage of tail intensity, tail length is also used. Data for tail intensity and tail length are represented in Figure 2.1. For these data, an nonnegligble set of dispersed observations are encountered and it was more pronounced for tail length. This may require attention in modeling with respect to the adequacy of the model to handle the dispersion present in the data. Further exploration is done to get an idea of the variation at rat- and slide-level. Figure 4.1 shows scatter plots of the average measurement at rat and slide levels after adjustment for the dose effect. Noticeable variability in the average score of the slides was observed, illustrating the importance of slide effect. Looking at the variability of the averages at slide level before and after adjusting for the rat effect, it can be seen that the variation shrinks more for tail length, implying that the rat effect could be more important for tail length as compared to tail intensity. This suggests the use of more elaborate models to formally check the importance of clustering and overdispersion.



(a) Tail Intensity



(b) Tail Length

Figure 4.1: *Comet Assay Study. Scatter plot of average (a) tail intensity and (b) tail length, after adjusting for the dose effect (Left): at rat level; (Middle): at slide level; (Right): at slide level, but adjusted for the rat effect.*

Standard methods to investigate the dose-response relationship of tail length and tail intensity, are based on first log-transforming the outcome to deal with the skewness of the outcome, and second, taking animal-averages of the log-transformed outcomes as a summary measure of the measurements in the animals. Thus, the hierarchical structure is completely ignored and the analysis is done using simple analysis of variance techniques. Sometimes, summary measures for the cells at slide-level are used instead of at the animal-level (Lovell and Omori 2008). The analysis is then performed using a mixed model, fitting the group as a fixed effect and animal as a random effect, and using the Kenward-Roger method for calculating degrees of

Table 4.1: *Comet Assay Study. Parameter estimates, standard errors and credible intervals from the conventional models for tail intensity.*

		Weibull		Analysis of variance	
Effect	Parameter	Estimate(s.e.)	95% C.I.	Estimate(s.e.)	95% C.I.
Veh.	β_0	-2.431(0.056)	[-2.54,-2.32]	0.234(0.052)	[0.13,0.34]
Low *vs.* veh.	β_1	-2.698(0.053)	[-2.80,-2.59]	3.351(0.073)	[3.21,3.49]
Med. *vs* veh.	β_2	-2.947(0.054)	[-3.05,-2.84]	3.527(0.074)	[3.38,3.67]
High *vs* veh.	β_3	-3.156(0.055)	[-3.27,-3.05]	3.693(0.074)	[3.55,3.84]
Pos. C.*vs* veh.	β_4	-1.711(0.060)	[-1.83,-1.59]	2.543(0.09)	[2.37,2.73]
Weibull shape	ρ	1.376(0.018)	[1.34,1.41]		

freedom (Kenward and Roger 1997). However, with this method, one loses a lot of information. Indeed, 150 cell observations are summarized by, for example, a single value. Such averaging effect may have a major impact on parameter estimation and corresponding inferences. Therefore, it is of paramount importance to deal with the full hierarchical structure using an appropriate probability distribution suggested in the literature. The results of the traditionally used analysis of variance and classical Weibull model for tail intensity are presented in Table 4.1, which will be compared to estimates from the proposed model in Section 4.5.

4.2 Hierarchical, Overdispersed, Non-Gaussian Continuous Outcomes

In this section, the model of interest for the comet assay data will be outlined. Because tail intensity and tail length are skewed, non-negative and continuous, which is similar to time-to-event data, an exponential or Weibull model appears appropriate. It is well known that the exponential distribution and gamma distribution are conjugate. The same holds for the Weibull distribution, when considered exponential in the outcome y_{ij}^{ρ}. These facts are reviewed in Molenberghs *et al* (2010). In particular, closed form expressions can be derived for the joint distribution, mean, variance, and higher-order moments.

Let us first assume that there is only one level of hierarchy in the data, e.g., the variability between animals. We then propose to use a combined model with a normal random effect to handle the hierarchy in the data and a conjugate random effect to account for overdisperion in the response. Using the Weibull distribution, this leads to

the Weibull-type combined model in (3.13)–(3.16) with b_i the animal-specific random effects to account for the clustering of observations and θ_{ij} the measurement-specific random effects to accommodate for overdispersion.

Next, let us propose an extension of the above model accounting for an extra level of hierarchy. Indeed, in the comet assay there are two sources of variation: one coming from the slide effect and one from the animal effect. The previous model can be extended by the use of three random effects of which one is the overdispersion effect. In addition, while typically a normal random effect is included in the linear predictor to account for the clustering, as in Molenberghs *et al* (2010), also a multiplicative factor using a multivariate gamma distribution can be used, similar to the multiplicative factor for the overdispersion random effect. For example, let us consider a model with a normally distributed random effect for the first hierarchy in the data and a gamma random effect for the second hierarchy in the data. In addition, we allow for the overdispersion in the model via another gamma-random effect. Let the outcome Y_{ijk} be the measurement for unit $k = 1, \ldots, n_{ij}$ of cluster $i = 1, \ldots, N$, sub-cluster $j = 1, \ldots, n_i$. The model can then be expressed as:

$$\begin{aligned} Y_{ijk}|b_i, b_{ij}, \theta_{ijk} &\sim \text{Weibull}(\rho, \lambda\theta_{ijk}b_{ij}e^{\eta_{ijk}}), \\ \eta_{ijk} &= \boldsymbol{x}'_{ij}\boldsymbol{\xi} + b_i, \\ \theta_{ijk} &\sim \text{Gamma}(\alpha_1, 1/\alpha_1), \\ b_{ij} &\sim \text{Gamma}(\alpha_2, 1/\alpha_2), \\ b_i &\sim \text{Normal}(0, D), \end{aligned}$$

leading to

$$f(y_{ijk}|\theta_{ijk}, b_i, b_{ij}) = \lambda\rho\theta_{ijk}b_{ij}y_{ijk}^{\rho-1}e^{\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+b_i}e^{-\lambda y_{ijk}^{\rho}\theta_{ijk}b_{ij}e^{\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+b_i}}, \quad (4.1)$$

$$f(\theta_{ijk}) = \frac{1}{\left(\frac{1}{\alpha_1}\right)^{\alpha_1}\Gamma(\alpha_1)}\theta_{ijk}^{\alpha_1-1}e^{-\alpha_1\theta_{ijk}}, \quad (4.2)$$

$$f(b_i) = \frac{1}{(2\pi d)^{1/2}}e^{-\frac{1}{2d}b_i^2}, \quad (4.3)$$

$$f(b_{ij}) = \frac{1}{\left(\frac{1}{\alpha_2}\right)^{\alpha_2}\Gamma(\alpha_2)}b_{ij}^{\alpha_2-1}e^{-\alpha_2 b_{ij}}. \quad (4.4)$$

The conditional mean, given the overdispersion and hierarchical random effects is:

$$E(y_{ijk}|\theta_{ijk}, b_i, b_{ij}) = \frac{\Gamma(\frac{1}{\rho}+1)}{\lambda\theta_{ijk}b_{ij}e^{\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+b_i}}. \tag{4.5}$$

Similarly, other models can be defined where either a gamma or a normal random effect is considered. This results in four different models: (a) Weibull- gamma(OD)-normal(RE1)-normal(RE2); (b) Weibull-gamma(OD)-normal(RE1)-gamma(RE2); (c) Weibull-gamma(OD)-gamma(RE1)-normal(RE2); and (d) Weibull-gamma(OD)-gamma(RE1)-gamma(RE2), where (.) explains what this random effect is considered for. OD refers to the overdispersion random effect, RE1 and RE2 refer to the first and second hierarchical random effect, respectively. As a result, a very flexible modeling framework is obtained for which model selection can easily be performed.

4.3 Bayesian Estimation Using MCMC

In the Bayesian framework, computation of the posterior probability is of main interest.The posterior probabilities are obtained by updating the likelihood with prior probabilities. For the Weibull-gamma(OD)-normal(RE1)-gamma(RE2) model for instance, combining the distribution of the outcome variable given in (7.6) with prior densities $f(\boldsymbol{\vartheta}, \theta_{ijk}, b_i, b_{ij})$, the posterior density is:

$$p(\boldsymbol{\vartheta}|y, x) \propto \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{n} f(y_{ijk}|\lambda_{ijk}) f(\boldsymbol{\vartheta}, \theta_{ijk}, b_i, b_{ij}), \tag{4.6}$$

where $\lambda_{ijk} = \theta_{ijk}b_{ij}\exp(\boldsymbol{x}'_{ijk}\boldsymbol{\xi} + b_i)$ and $\boldsymbol{\vartheta}$ is a group of parameters $(\boldsymbol{\xi}, \rho)$.

Sampling was done with two chains and dispersed initial values were given for all parameters in the two chains. 150,000 samples were drawn from each chain and the first 100,000 samples were discarded. To ensure the samples are drawn from the target posterior density, convergence was checked by comparing the between- and within-chain variation for each parameter in the simulated samples.

Non-informative or weak priors were used for all the parameters of interest: $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4 \sim N(0, 10^6)$; $\sigma_r^2 \sim IG(0.1, 0.001)$; $\rho \sim \exp(0.01)$; $\alpha_\theta \sim \text{Gamma}(2, 2)$; and $\alpha_s \sim \text{Gamma}(0.1, 0.1)$, where IG is the Inverse Gamma distribution. The mean was reported as point estimate for each parameter, together with the 95 percent credible interval that ranges between the 2.5 and 97.5 percent quantiles. Note that, while

the α value for the gamma distribution will result in a relatively informative prior, sensitivity of the prior has to be checked. In general, for large datasets, varying this value has little or no impact on the conclusion. For relatively small sets of data, however, caution is needed.

4.4 Illustration of Overdispersion and of the Clustering Effect

Overdispersion, in which the variability in the data is beyond the variance of the model considered, occurs quite often in practice. This is basically because of the restricted relationship between mean and variance functions. The extra-variability could be due to some unaccounted covariates/factors, heterogeneous population, clustering effect and many others. The extra unaccounted variability can be accounted for by the use of mixture models or using an overdispersion parameter. In this case, a continuous overdispersion random effect is used. We now illustrate how the effect of the overdispersion random effect and the clustering random effect extend the Weibull model to accommodate more dispersed data. This is done by comparing the marginalized densities corresponding to different models. Because no analytical expressions are available for all these models, a large dataset is simulated from the models, and a density plot is made from the data. we simulate data from models that assume both overdispersion and clustering, model that consider either overdispersion or clustering and a model that considers neither of them. Let us first see the effect of overdispersion alone. The data are generated from: (1) Weibull Model: Weibull$(\rho, 1)$ and (2) Weibull-gamma Model: Weibull(ρ, θ). In the Weibull-gamma model, we assume different choices for the Gamma$(\alpha, 1)$ distribution (different choices of α parameters) in order to see the effect of overdispersion. The result is given in the upper panel of Figure 4.2. Inclusion of the gamma random effect, results in ticker right-tails, an important characteristic of overdispersion. Similarly, to have an idea of the effect of clustering alone and the effect of overdispersion combined with clustering, we generate data from the following set: (1) Weibull model: Weibull$(\rho, 1)$; (2) Weibull-normal model: Weibull(ρ, e^{b_i}); and (3) Weibull-gamma-normal model: Weibull$(\rho, \theta e^{b_i})$, where $\rho = 1.4$, $b_i \sim$ Normal$(0, 1)$, $i = 1, \ldots, 20$, and $\theta \sim$ Gamma$(\alpha, 1)$.

The density plot of the data is presented in the lower panel of Figure 4.2. Introducing clustering leads to more dispersion in the data. With the inclusion of both the overdispersion and clustering random effect, more variability is seen with the degree

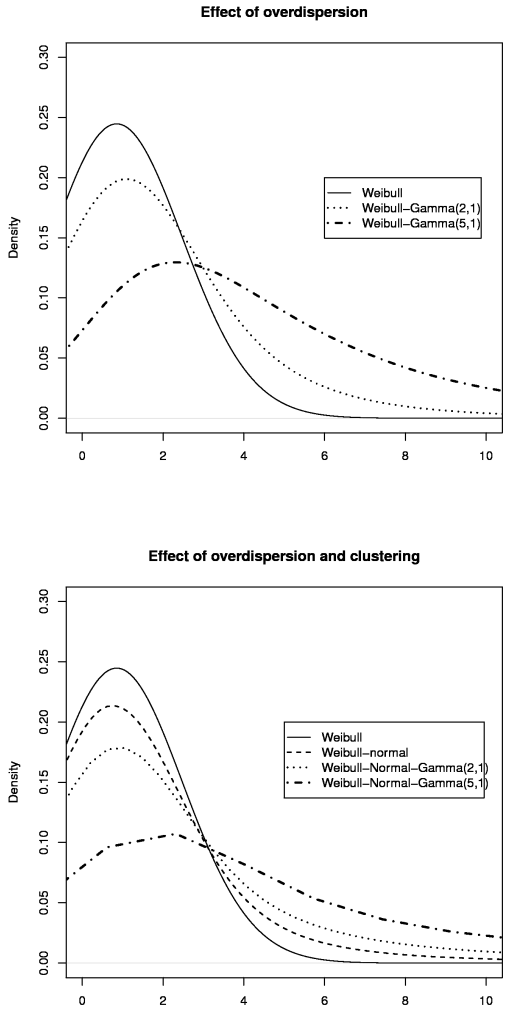


Figure 4.2: *Illustration of the effect of clustering and overdispersion.*

of dispersion depending on the α parameter. If this is the case, analyzing more dispersed data using the traditional model, which does not address overdispersion and clustering, may leave a lot of variability unexplained and it can affect the precision of the estimates.

4.5 Application to the Comet Assay Data

The primary goal is to assess the toxicity of 1,2-Dimethylhydrazine dihydrochloride at different dose levels. The data described in Chapter 2 are analyzed taking the multilevel hierarchical nature into account. Cells coming from the same rat could be more alike due to biological reasons, implying clustering at the animal level. Moreover the fact that cells are grouped into three slides could pose some sort of clustering due to uncontrolled differences in external factors such as the amount of gel being used.

In addition to the hierarchical structure, the skewed nature of the outcome variable adds complexity. Tail intensity and tail length are non-negative continuous outcomes. In the literature, a number of probability distributions were proposed for modelling the distribution. These include the Weibull, exponential, logistic, normal, log-normal, and log-logistic distributions (Lovell and Omori 2008). Ejchart and Sadlej-Sosnowska (2003) found that Weibull was the best distribution for such data. In our analyses, also the Weibull distribution was assumed. As explained in the previous section, a gamma random effect is used for the overdispersion and both normal and gamma random effects are used to explain the hierarchical structure. As described in Section 4, the normal random effect is included in the linear predictor and the gamma random effect is included as multiplicative effect together with the overdispersion factor. This creates a wide choice of models to choose from. Table 4.2 presents an overview of the models considered.

Model 1 is the traditional Weibull model that ignores the hierarchical nature as well as overdispersion. Model 4 considers the overdispersion but not the hierarchical nature. Models 2, 3, 7, and 8 consider one random effect (rat or slide) and ignore the over-dispersion and the other random effect, the classical gamma frailty model being part of it. Models 5, 6, 9, and 10 consider the overdispersion and one random effect but ignore the second one. Models 11–14 consider the correct hierarchical structure but ignore the overdispersion. The last four models account for both the hierarchical nature and overdispersion.Let Y_{ijk} represent the tail intensity or tail length measured for a k^{th} cell ($k = 1, \ldots, 50$) from rat $i = 1, \ldots, 27$ in slide $j = 1, 2, 3$. If we consider

Table 4.2: *Comet Assay Study. Overview of models considered with DIC for tail intensity(TI)and tail length(TL).*

	Distribution for							
	Response	Overdispersion	RE1(rat)		RE2(slide)			
Model	Weibull	Gamma	Normal	Gamma	Normal	Gamma	DIC(TI)	DIC(TL)
1	√						33869.6	30878.8
2	√		√				33823.9	30421.6
3	√			√			33823.5	30420.2
4	√	√					33895.6	27378.5
5	√	√	√				33853.7	26901.6
6	√	√		√			33852.5	26883.0
7	√				√		**33728.9**	29622.6
8	√					√	**33728.5**	29620.8
9	√	√			√		33760.7	26386.9
10	√	√				√	33760.6	26377.0
11	√		√		√		**33728.7**	29623.4
12	√		√			√	**33728.6**	29619.5
13	√			√	√		**33730.3**	29631.1
14	√			√		√	**33729.7**	29605.2
15	√	√	√		√		33761.6	26374.4
16	√	√	√			√	33760.5	26333.1
17	√	√		√	√		33760.6	26338.0
18	√	√		√		√	33758.6	**26209.6**

the Weibull-Gamma(OD)-Normal(RE1)-Normal(RE2) model, for instance, the λ_{ijk} will be modelled as:

$$\lambda_{ijk} = \theta_{ijk} \exp(\beta_0 + \beta_1 L_{ijk} + \beta_2 M_{ijk} + \beta_3 H_{ijk} + \beta_4 PC_i + r_i + s_{ij}), \tag{4.7}$$

with measurement-specific random effects $\theta_{ijk} \sim \text{Gamma}(\alpha_1, 1/\alpha_1)$, rat-specific random effects $r_i \sim N(0, d_1)$ and slide-specific random effects $s_{ij} \sim N(0, d_2)$. Further, L_{ijk} is the indicator variable whether rat i is given a Low dose (1 if it is given low dose; 0 otherwise). Similarly, M_{ijk}, H_{ijk}, PC_{ijk} are the indicator variables for medium dose, high dose, or positive control, respectively. Similarly, the Weibull-gamma(OD)-normal(RE1)-gamma(RE2) is parameterized as:

$$\lambda_{ijk} = \theta_{ijk} * s_{ij} * \exp(\beta_0 + \beta_1 * L_{ijk} + \beta_2 * M_{ijk} + \beta_3 * H_{ijk} + \beta_4 * PC_{ijk} + r_i), \tag{4.8}$$

with now $\theta_{ijk} \sim \text{Gamma}(\alpha_1, 1/\alpha_1)$, $r_i \sim N(0, d)$ and $s_{ij} \sim \text{Gamma}(\alpha_2, 1/\alpha_2)$. The fixed effect β_0 denotes the control (vehicle) effect. The parameters β_1 to β_4 are the contrasts of interest that represent the effect of low dose, medium dose, high dose, and positive control versus vehicle. All other models follow similarly. The R2winbugs code for Models 8 and 16 is given in Appendix A.

The next issue is model comparison. In situations where non-informative priors are used or when where huge amounts of data are available, the data overwhelm the choice of the prior and Bayesian estimates are equivalent with maximum likelihood estimates. In such a case, the likelihood ratio can be used to formally test hypotheses and compare nested models. In this case, the deviance information criterion (DIC) is used as a model comparison tool. It penalizes for the complexity of the model as explained in Section 3.2.3.1. Based on the deviance, which favors complex models, Weibull-gamma(OD)-gamma(RE1)-gamma(RE2) was to be preferred. Note that the DIC is subject to random variability and hence differences in value by 2–4 should not be regarded as evidence for a difference. Therefore, Models 7 and 8 on the one hand and Models 11–14 on the other can be regarded as roughly equivalent for tail intensity. For the outcome tail length, model 18 is the best model. Note that the penalty term, measuring the complexity of the model, for the models with overdispersion was large. Looking at models 8 and 12, Model 12 did not outperform and this was not unexpected from the exploratory data analysis for tail intensity, the variability at the slide level did not reduce much after removing the rat effect (Figure 4.1).

The parameter estimates and 95 percent credible intervals for Models 8 and 12 for the tail intensity are presented in Table 4.3. Models 12 and 8 are nested models, where Model 12 is an extension of model 8 by the inclusion of the normal random effect for the animal level. The parameter estimates from both models are very similar and we notice that the standard errors from Model 12 are consistently and slightly higher as opposed to the ones obtained with Model 8. This is in line with the expectation that, with exclusion of one hierarchical level, the effective degrees of freedom is usually overestimated which results in underestimation of the standard errors.

The parameter estimates from the model with overdispersion only (Model 4) had higher standard errors compared to the estimates from a classical Weibull model. In addition, they were lower compared to that of Weibull-gamma(RE2), the preferred model. The estimates from models with both overdispersion and clustering have higher standard errors compared with models with either overdispersion or clustering. The 95% credible interval for ρ did not include 1, which conveys that the Weibull distribution is more plausible than the exponential. The 95% credible interval for the

Table 4.3: *Comet Assay Study. Parameter estimates, standard errors, and credible intervals of the regression coefficients in Model 8 (Weibull-gamma(RE2) model), and Model 12 (Weibull-normal(RE1)-gamma(RE2) model) for tail intensity.*

		Models 8		Models 12	
Effect	Par.	Est.(s.e.)	95% C.I.	Est.(s.e.)	95% C.I.
Veh.	β_0	-2.419(0.079)	[-2.57,-2.26]	-2.427(0.085)	[-2.59,-2.25]
Low *vs* veh.	β_1	-2.854(0.097)	[-3.04,-2.66]	-2.850(0.104)	[-3.06,-2.65]
Med.*vs* veh.	β_2	-3.092(0.098)	[-3.29,-2.90]	-3.088(0.106)	[-3.30,-2.88]
High *vs* veh.	β_3	-3.317(0.098)	[-3.51,-3.12]	-3.312(0.107)	[-3.53,-3.11]
Pos. C. *vs* veh.	β_4	-1.829(0.115)	[-2.05,-1.60]	-1.826(0.124)	[-2.07,-1.58]
Weibull shape	ρ	1.420(0.019)	[1.38,1.46]	1.419 (0.019)	[1.38,1.46]
Precision of RE1	$\frac{1}{d}$			114.2(79.29)	[28.60,331.61]
RE2 parameter	α_2	18.33(4.036)	[11.68,27.3]	19.99(4.493)	[12.08,29.54]

regression parameters describing treatment contrasts of interest did not include zero indicating toxicity of the chemical at all dose levels. This same final conclusion was reached by all models. However, the credible intervals were affected by the choice of the model.

As explained in Section 4.1, conventional analyses transform the tail intensities using logarithmic transformations. The mean of the transformed responses is then used as a summary measure for each rat. The hierarchical nature of the data is thus completely ignored and a simple analysis of variance is used to test whether there is a dose effect. Comparing this conventional model (Table 4.1) to our preferred model would be rather difficult since as we are using different responses and different type of models. We can, however, compare this with an equivalent model from our set of proposed models which completely ignores the hierarchical structure, but which uses the appropriate distribution and all the available information, namely the classical Weibull model. Upon comparison of the classical Weibull model with Model 8, the parameters of interest are highly significant in both cases. Yet, the standard errors, likewise the credible intervals of Model 8 are twice that of the classical Weibull model. While not the case in this example because of the high toxicity of the compound of interest, this suggests that ignoring the hierarchical structure and overdispersion could have major influence on the final conclusion. Significant estimates in the classical Weibull model may be insignificant in Model 8. In other words, a compound might be erroneously declared toxic.

Based on the analysis for tail intensity, more elaborate models did not outperform

Table 4.4: *Comet Assay Study. Parameter estimates, standard errors, and credible intervals of the regression coefficients in (1) the Weibull-gamma(OD)-gamma(RE1)-gamma(RE2) model, Model 18, (2) the Weibull-gamma(OD)-normal(RE1)-gamma(RE2) model, Model 16 for Tail Length.*

		Model 18		Model 16	
Effect	Parameter	Est.(s.e.)	95% C.I.	Est.(s.e.)	95% C.I.
Veh.	β_0	-30.44(0.66)	[-31.74,-29.12]	-30.54(0.80)	[-32.01,-28.97]
Low *vs* veh.	β_1	-11.99(0.50)	[-12.95,-11.01]	-12.02(0.52)	[-13.05,-11.06]
Med. *vs* veh.	β_2	-12.14(0.51)	[-13.1,-11.12]	-12.19(0.53)	[-13.27,-11.23]
High *vs* veh.	β_3	-12.57(0.49)	[-13.54,-11.58]	-12.63(0.54)	[-13.75,-11.64]
Pos. C. *vs* veh.	β_4	-9.75(0.55)	[-10.84,-8.65]	-9.75(0.56)	[-10.88,-8.68]
Weibull shape	ρ	10.71(0.22)	[10.26,11.13]	10.71(0.2727)	[10.17,11.22]
Precision of RE1	$\frac{1}{d}$			32.14(168.10)	[1.27,323.2]
OD parameter	α_1	0.89(0.04)	[0.82,0.98]	0.89(0.049)	[0.81,0.99]
RE1 parameter	α_2	4.60(3.18)	[1.53,12.67]		
RE2 parameter	α_3	1.61(0.30)	[1.10,2.25]	1.58(0.31)	[1.05,2.28]

(not much improvement in terms of DIC). However, this was not the case for the second response, tail length. Based on the DIC, the most complicated model has the best fit, showing the importance of the hierarchical structure as well as overdispersion, as shown in Table 4.2. Models with one hierarchical random effect were better fitting as compared to the classical Weibull model. Models with two random effect improved the fit further, and models with the complete hierarchical structure and overdispersion random effect appear to be best. Further, notice that the model with only the overdispersion random effect is better fitting than models with only the hierarchical structure, showing the importance of the overdispersion relative to the hierarchical structure.

Note the effects of the model on the parameter estimates (Table 4.5). When only one hierarchical structure (one random effect) is added to the classical Weibull model, the point estimates were slightly higher and the standard error for the contrast of interest was approximately four times larger. Smaller DIC for models with the second random effect (slide) showed the importance of slide effect in contrast to rat. Extending to two random effects, the standard error slightly increased further. The inclusion of an overdispersion random effect had a very important impact on the estimate (approximately 3 times) and standard error (four times) in contrast with the classical Weibull model. With the inclusion of one hierarchical random effect to the overdis-

Table 4.5: *Comet Assay Study. Parameter estimates, and standard errors of the regression coefficients in the Weibull model (Model 1), the Weibull-gamma(OD) (Model 4), Weibull-gamma(RE2) (Model 8), and the Weibull-normal(RE1)-gamma(RE2) (Model 12) for Tail Length.*

		Model 1	Model 4	Model 8	Model 12
Effect	Parameter	Est.(s.e.)	Est.(s.e.)	Est.(s.e.)	Est.(s.e.)
Veh.	β_0	-12.76(0.15)	-27.36(0.60)	-15.26(0.23)	-15.26(0.25)
Low *vs* veh.	β_1	-3.55(0.05)	-10.58(0.26)	-4.80(0.22)	-4.79(0.25)
Med. *vs* veh.	β_2	-3.65(0.05)	-10.76(0.26)	-4.90(0.22)	-4.89(0.25)
High *vs* veh.	β_3	-3.85(0.06)	-11.13(0.27)	-5.10(0.22)	-5.10(0.25)
Pos. c. *versus* veh.	β_4	-2.70(0.06)	-8.55(0.22)	-3.81(0.26)	-3.79(0.30)
Weibull shape	ρ	4.01(0.04)	9.48(0.22)	4.97(0.06)	4.96(0.06)
Precision of RE1	$\frac{1}{d}$				45.83(54.60)
OD parameter	α_1		0.86(0.04)		
RE2 parameter	α_3			2.79(0.45)	3.03(0.54)

persion, the standard error was doubled. Models with complete hierarchical structure and overdispersion yielded a slightly different estimate compared to the estimate from a model with overdispersion alone and a changing estimate (approximately 2.5 times) in contrast to the estimate of the corresponding models with two hierarchical random effects but no overdispersion; the standard error was double in contrast to both models. Generally, for tail length, we did not reach a different conclusion, due to high toxicity of the compound; however, inclusion of the hierarchical structure and overdispersion random effect had severe impact on the magnitude, standard errors as well as the credible intervals. Results for the more elaborate models with complete hierarchical structure and overdispersion are given in Table 4.4 and the results for the classical Weibull model, a model with overdispersion alone, and models with a single and two hierarchical random effects are given in Table 4.5.

4.6 Concluding Remarks

In this chapter, a flexible modeling framework for the comet assay data using a Bayesian hierarchical model. It takes not only the complete hierarchical nature but also the appropriate non-Gaussian probability distribution for the response into account. It further includes a possible overdispersion that may exist in the data. Both normal and gamma random effects can be considered to account for clustering in the same framework, the more conventional models with either the overdispersion, or just one hierarchical random effect being submodels.

The method was applied to the comet assay data gathered to assess the toxicity of 1,2-Dimethylhydrazine dihydrochloride at different dose levels. For this particular dataset, a Weibull-gamma(RE2) model seemed adequate for tail intensity, whereas a Weibull-gamma(OD)-gamma(RE1)-gamma(RE2) was better fit for tail length. A comparison of these analysis with the conventional approach, which ignores the overdispersion and the hierarchy in the data, revealed that both models led to the same qualitative conclusion of severe toxicity of the compound at all dose levels. This notwithstanding, estimates, standard errors, and credibility intervals were severely affected, underscoring the risk of using models that are too simple. In general, proper models encompassing at the same time the hierarchical nature in the data, combined with overdispersion effects, need to be adopted. In this case, the use of the overdispersion and hierarchical structure improved the fit for one response. Furthermore, even when the more elaborate model does not provide a substantially improved fit, nor alters the inferences drawn, the development is still very useful because it provides further confidence, by way of model specification assessment, on the quality of the purported model.

Chapter 5

Joint Modeling of Hierarchically Clustered and Overdispersed Outcomes for Comet Assay Data

In the previous chapter, we have dealt with univariate response. However, it is not always appropriate to do analyses on a single endpoint. Multivariate longitudinal or clustered data are also commonly encountered in clinical trials and toxicological studies. Typically, there is no single standard endpoint to assess the toxicity or efficacy of the compound of interest, but multiple endpoints, the so-called co-primary endpoints, are available to assess the toxic effects or the activity of the compound. In a comet assay, for instance, different outcomes (Lovell and Omori 2008, Wiklund and Agurell 2003) are used to assess the DNA damage of a cell as a result of an exposure: the tail length, tail intensity, and tail moment. These outcomes will formally be introduced in the next section. Most often, the tail length and tail intensity are used. Typically, univariate analyses are conducted to asses the treatment effect on each endpoint separately, leading to as many conclusions as there are endpoints regarding the same treatment effect. In particular, for the comet assay, one tends to focus primarily on tail intensity because of its discriminative power. Ideally though, one would reach a conclusion on the overall effect using all outcomes simultaneously,

necessitating a joint analysis. An added value of joint modeling is that inferences can be drawn about the association between outcomes as well.

Various modeling approaches for specifying a joint distribution are possible (Fitzmaurice *et al.*, 2009, Ch. 14; Fieuws and Verbeke, 2004). First, this can be effectuated by *specifying the full multivariate distribution* of the outcomes. This allows for drawing marginal inferences regarding the characteristics of the individual outcomes, but it requires many parameters and while the multivariate Gaussian distribution is well-known, there are many distributions for which no commonly accepted multivariate distribution is available. Second, i can be done by the use of *conditional models* where the joint distribution is expressed as the product of the conditional distribution of the first outcome conditional on the second outcome and the marginal distribution of the second outcome. However, factorization can be done in many ways, leading to different results, and it requires the specification of many parameters. Third, *shared-parameter models* can be entertained, where a pair of outcomes are associated by using a common latent variable, e.g., a common random effect. This is a simple but very strong assumption about the association between outcomes. Fourth, one can relax the latter assumption by using *multivariate random effects*, in which the two outcomes are associated via separate correlated random effects. This is more flexible than shared-parameter models, but might still fail to fully capture the association structure and/or the variance function. Fifth, dimension reduction using principal components can be used, upon which the principal components are subjected to univariate analysis. While simple, the resulting inferences may not be about the parameters of direct scientific interest. In this chapter, we focus on a flexible multivariate random effects approach. This work has been published in Ghebretinsae *et al* (2012).

The joint model for two hierarchical, overdispersed non-Gaussian outcomes is outlined in Section 5.1, and characteristics of the models are derived. It is then applied to the comet data in Section 5.2.

5.1 Joint Model for Two Hierarchical, Overdispersed Positive Outcomes

In this section, a joint model for hierarchical, overdispersed positive outcomes is proposed. First, the setting of a single hierarchical, overdispersed outcome is introduced, which is then extended to the multivariate setting.

5.1.1 Univariate Analysis

As explained in the previous chapters, because the primary outcomes, tail intensity and tail length, are skewed, non-negative and continuous, which is similar to many time-to-event data (Duchateau and Janssen 2007), an exponential or Weibull distribution is a natural choice. Here, we account for one level in the hierarchy of the data, namely the variability between slides. As proposed by Ghebretinsae *et al* (2011), we use a combined Weibull model with normal random effects to handle the hierarchy in the data and a gamma conjugate random effect to account for overdispersion in the response. This model falls into the model family as proposed by Molenberghs *et al.* (2010).

Let Y_{ij} be the j^{th} cell of subject i measured for tail length or tail intensity, grouped in to $\boldsymbol{Y}_i$. The Weibull-type combined model (Weibull-gamma-normal model) in (3.13)–(3.16) is considered. $\boldsymbol{b_i}$ is the zero-mean normally-distributed slide-specific random effects, with variance-covariance D, to account for the clustering of observations and θ_{ij} is the gamma-distributed measurement-specific random effects to accommodate for overdispersion. Further, λ and ρ are Weibull parameters, and α_j and β_j are gamma parameters. Here, η_{ij} is a linear predictor, with fixed-effects parameter $\boldsymbol{\xi}$ and design vectors $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ for the fixed effects and random effects, respectively.

5.1.2 Joint Analysis

The proposed joint model for tail length and tail intensity assumes a Weibull-gamma-normal model for both endpoints. The endpoints are associated by the use of bivariate normal random effects for the two endpoints, instead of the use of two separate (univariate) random effects, which we will call the Weibull-Gamma-Multivariate Normal model.

Let Y_{1ij} and Y_{2ij} be the j^{th} measurements of subject i for the two outcomes, tail length and tail intensity respectively. With notation similar to the above, the linear part for the two responses are assumed to be:

$$\begin{aligned} \eta_{1ij} &= \boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + b_{1i}, \\ \eta_{2ij} &= \boldsymbol{x}'_{2ij}\boldsymbol{\xi}_2 + b_{2i}, \end{aligned}$$

with $\boldsymbol{x}_{1ij}$ and $\boldsymbol{x}_{2ij}$ design matrices, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ vectors of unknown fixed-effect parameters, and b_{1i} and b_{2i} the cluster-specific random intercepts for the first and second outcomes, respectively. These two random effects are assumed bivariate normally

distributed:

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_1^2 & rd_1d_2 \\ rd_1d_2 & d_2^2 \end{pmatrix}\right], \tag{5.1}$$

with d_{11}^2 and d_{22}^2 the variances of the random intercepts and r the correlation between them. The association between the two endpoints is induced via the parameter r. More details on this are given in the next section. Conditionally on the normally distributed random effects (b_{1i} and b_{2i}), it is assumed that the two outcomes are independent. Testing for treatment effect based on both endpoints simultaneously is conveniently done by way of a likelihood ratio test for the treatment effect parameters in both endpoints combined.

In terms of estimation, we opt for maximum likelihood using partial marginalization. This implies that the gamma random effects are analytically integrated out from the likelihood, while numerical integration, as implemented in the SAS procedure NLMIXED, is invoked to marginalize over the normally distributed random effects. The code is given in the Appendix B.1.

5.1.3 Correlation Between Both Responses

The association between both outcomes is captured via the bivariate normal random effects. However, the correlation between the two random effects is not necessarily equal to the correlation between the two responses. Furthermore, a significant correlation at the cluster level does not necessarily imply a significant correlation between the two responses taken from the same cell. In this section, it is established how the correlation between the outcomes is related with the correlation between the random effects.

The correlation between two measurements from the same subject for a single response, also called the intraclass correlation (ICC), is equal to:

$$\text{Corr}(Y_{\ell ij}, Y_{\ell ik}) = \left[e^{\frac{d_\ell^2}{\rho_\ell^2}} - 1\right] \Bigg/ \left[\frac{2\rho_\ell B(\alpha_\ell - \frac{2}{\rho_\ell}, \frac{2}{\rho_\ell}) e^{\frac{d_\ell^2}{\rho_\ell{}^2}}}{B(\alpha_\ell - \frac{1}{\rho_\ell}, \frac{1}{\rho_\ell})^2} - 1\right]$$

with $\ell = 1, 2$, ρ_ℓ the shape parameter of the Weibull distribution, d_ℓ^2 the random-effects variance (equal to d_1^2 or d_2^2 for tail length and tail intensity, respectively), and α_ℓ the shape parameter of the Gamma random effects distribution. $B(\cdot, \cdot)$ is

the beta function. A large value for the shape parameter (α_ℓ) indicates a small amount of overdispersion, which in the limit reduces to the Weibull-Normal model for a univariate outcome. In this case, the intraclass correlation reduces to:

$$\text{Corr}(Y_{\ell ij}, Y_{\ell ik}) = \left[e^{\frac{d_\ell^2}{\rho_\ell^2}} - 1\right] \Bigg/ \left[\frac{2\rho_\ell \Gamma(\frac{2}{\rho_\ell}) e^{\frac{d_\ell^2}{\rho_\ell{}^2}}}{\Gamma(\frac{1}{\rho_\ell})^2} - 1\right],$$

with $\Gamma(\cdot)$ the gamma function.

On the other hand, the correlation between the two outcomes (tail length and tail intensity) of the same cell is given by the following expression

$$\begin{aligned}
\text{Corr}(Y_{1ij}, Y_{2ij}) &= \left(e^{\frac{r d_1 d_2}{\rho_1 \rho_2}} - 1\right) \\
&\quad \times \frac{B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1}\right)}{\left[2\rho_1 B\left(\alpha_1 - \frac{2}{\rho_1}, \frac{2}{\rho_1}\right) e^{\frac{d_1^2}{\rho_1{}^2}} - B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1}\right)^2\right]^{1/2}} \\
&\quad \times \frac{B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)}{\left[2\rho_2 B\left(\alpha_2 - \frac{2}{\rho_2}, \frac{2}{\rho_2}\right) e^{\frac{d_2^2}{\rho_2{}^2}} - B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)^2\right]^{1/2}} \\
&= \frac{\left(e^{\frac{r d_1 d_2}{\rho_1 \rho_2}} - 1\right)}{\left(e^{\frac{d_1^2}{\rho_1^2}} - 1\right)^{\frac{1}{2}} \left(e^{\frac{d_2^2}{\rho_2^2}} - 1\right)^{\frac{1}{2}}} \sqrt{\text{ICC}_1} \sqrt{\text{ICC}_2}, \qquad (5.2)
\end{aligned}$$

where ICC_1 and ICC_2 are the intracluster/class correlation for responses 1 and 2, respectively. The correlation between the two endpoints is proportional to the correlation between the two random effects, with the same sign. So, when two random effects are positively or negatively correlated, the correlation between endpoints follows accordingly and when the correlation between the two random effects is zero, then the correlation between the two endpoints is zero as well. In other words, the correlation is induced entirely by the correlation between the two random effects. This correlation also depends on the Weibull shape parameters ρ_1 and ρ_2.

For a joint model based on two linear mixed models, the bivariate correlation between the two endpoints is given by $\text{Corr}(Y_{1ij}, Y_{2ij}) = r\sqrt{\text{ICC}_1}\sqrt{\text{ICC}_2}$ (Fitzmaurice *et al.*, 2009, Ch. 14). It is by definition smaller than or equal to the correlation between the two random intercepts. Only when both intra-class correlations are 1, equality

holds. However, it is not straightforward in this case. Details on the calculations are given in Appendix B.2.

5.2 Application to the Comet Data

5.2.1 Univariate Analyses

Univariate analyses for tail intensity and tail length are performed separately for the comet assay data (Figure 5.1). The endpoints are analyzed both with and without overdispersion, using the Weibull-gamma-normal and Weibull-normal models, respectively. Summary results are presented in Tables 5.1 and 5.2. For tail intensity, inclusion of the overdispersion random effect neither improved the likelihood, nor affected the estimates and precision of the estimate. On the other hand, for tail length, inclusion of the overdispersion random effect greatly improved the likelihood and also affected the parameter estimation and precision. If we consider the contrast between the low and high dose group for tail length, for instance, the p-value was 0.1358 based on the model without overdispersion and 0.0543 with overdispersion. For both endpoints, there is a major effect of the compound as compared to the vehicle group. However, the conclusion for the contrasts between the three dose level is different based on both responses. Based on tail intensity, there was a significant difference among the dose levels.

A conventional significance test for α would test the null hypothesis $H_0 : \alpha = 0$. However, this does not correspond to the absence of overdispersion. Rather, overdispersion vanishes as α approaches infinity.

5.2.2 Analysis Based on a Combined Endpoint

It is often desirable to opt for a summary analysis of both endpoints, at least to avoid multiple and perhaps conflicting inferences from the univariate analyses. To this end, define tail moment as the product of the mean distance of migration in the tail with the amount of DNA in the tail (intensity). Although not directly the product of the two responses, it indirectly combines information from both endpoints. Also here, inclusion of overdispersion improved the fit and had impact on the parameter estimates as well as on the standard errors; see Table 5.3. Remember that the contrasts between low, medium and high doses was significant based on the univariate analysis using tail intensity, but not using the endpoint tail length. Using tail moment, none of the

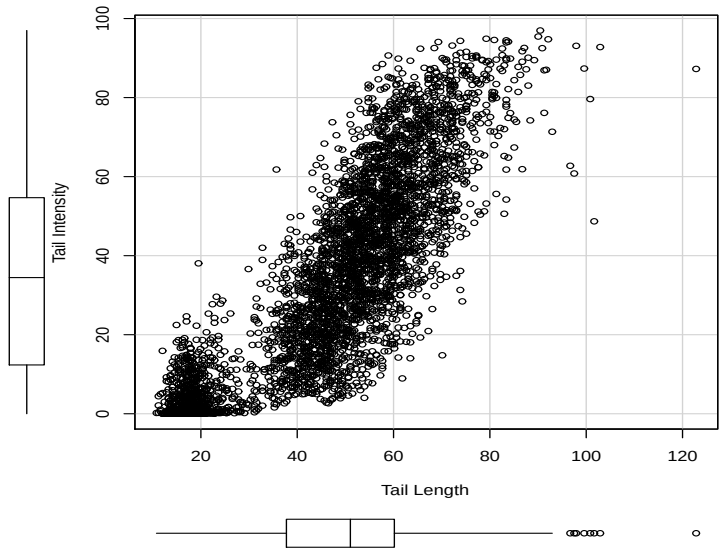


Figure 5.1: *Comet Assay data. Scatter plot and box plots of the tail length versus tail intensity*

Table 5.1: *Comet Assay Study. Parameter estimates, and standard errors for the regression coefficients in (1) the Weibull-gamma-normal model, (2) the Weibull-normal model in the analysis for tail length*

		Weibull-gamma-normal		Weibull-normal	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value
Veh.	β_0	-30.9295(0.7264)	0.0001	-15.6378(0.2517)	0.0001
Low *vs.* veh.	β_1	-11.9378(0.4445)	0.0001	-4.4965(0.2243)	0.0001
Med.*vs.* veh.	β_2	-12.1552(0.4472)	0.0001	-4.5998(0.2245)	0.0001
High *vs.* veh.	β_3	-12.6026(0.4525)	0.0001	-4.8290(0.2251)	0.0001
Pos. C.*vs.* veh.	β_4	-9.6419(0.4762)	0.0001	-3.4808(0.2718)	0.0001
Low *vs.* Med.	β_5	-0.2174(0.3398)	0.5241	-0.1033(0.2206)	0.6410
Low *vs.* High	β_6	-0.6648(0.3403)	**0.0543**	-0.3325(0.2206)	**0.1358**
Med. *vs.* High	β_7	-0.4474(0.3402)	0.1923	-0.2292(0.2206)	0.3019
Weibull Par.	ρ	10.7072(0.2474)	0.0001	4.9585(0.0580)	0.0001
s.d. of RE	$\sqrt{d}$	0.9881(0.08592)	0.0001	0.6464(0.0543)	0.0001
OD par.	α	0.8932(0.0463)	0.0001	–	–
-2 loglik.		28069		29793	

contrasts are significant. This shows that some effects might be lost by summarizing

Table 5.2: *Comet Assay Study. Parameter estimates, and standard errors for the regression coefficients in (1) the Weibull-gamma-normal model, (2) the Weibull-normal model in the analysis for tail intensity*

		Weibull-gamma-normal		Weibull-normal	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value
Veh.	β_0	-2.4628(0.0774)	0.0001	-2.4628(0.0774)	0.0001
Low *vs.* veh.	β_1	-2.8125(0.0911)	0.0001	-2.8126(0.0911)	0.0001
Med.*vs.* veh.	β_2	-3.0565(0.0920)	0.0001	-3.0566(0.0920)	0.0001
High *vs.* veh.	β_3	-3.2777(0.0929)	0.0001	-3.2778(0.0929)	0.0001
Pos. C. *vs.* veh.	β_4	-1.7941(0.1079)	0.0001	-1.7941(0.1078)	0.0001
Low *vs.* Med.	β_5	-0.2440(0.0874)	0.0065	-0.2440(0.0874)	0.0065
Low *vs.* High	β_6	-0.4652(0.0875)	0.0001	-0.4652(0.0875)	0.0001
Med. *vs.* High	β_7	-0.2212(0.0874)	0.0133	-0.2212(0.0874)	0.0133
Weibull Par.	ρ	1.4158(0.0189)	0.0001	1.4158(0.0189)	0.0001
s.d. of RE	$\sqrt{d}$	0.2201(0.0248)	0.0001	0.2201(0.0248)	0.0001
log OD par.	$\log(\alpha)$	13.9715(2.0370)	0.0001	–	–
-2 loglik.		33769		33769	

Table 5.3: *Comet Assay Study. Parameter estimates, and standard errors for the regression coefficients in (1) the Weibull-gamma-normal model, (2) the Weibull-normal model in the analysis for tail moment*

		Weibull-Gamma-Normal		Weibull-Normal	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value
Veh.	β_0	1.0946(0.1023)	0.0001	0.8294(0.0781)	0.0001
Low *vs.* veh.	β_1	-3.8954(0.1515)	0.0001	-3.4995(0.1159)	0.0001
Med.*vs.* veh.	β_2	-4.2326(0.1558)	0.0001	-3.8116(0.1171)	0.0001
High *vs.* veh.	β_3	-4.5767(0.1599)	0.0001	-4.1341(0.1186)	0.0001
Pos. C. *vs.* veh.	β_4	-2.2365(0.1594)	0.0001	-1.9365(0.1360)	0.0001
Low *vs.* Med.	β_5	-0.3372(0.1174)	0.0052	-0.3121(0.1104)	0.0059
Low *vs.* High	β_6	-0.6813(0.1180)	0.0001	-0.6346(0.1106)	0.0001
Med. *vs.* High	β_7	-0.3441(0.1174)	0.0044	-0.3224(0.1104)	0.0046
Weibull Par.	ρ	1.3199(0.0237)	0.0001	1.2429(0.0159)	0.0001
s.d. of RE	$\sqrt{d}$	0.3174(0.0316)	0.0001	0.2990(0.0291)	0.0001
log OD par.	$\log(\alpha)$	10.9408(2.6443)	0.0001	–	–
-2 loglik.		19980		20004	

two endpoints by a single endpoint.

5.2.3 Joint Analysis

The univariate analyses on the two endpoints (Section 5.2.1) lead to multiple inferences. The univariate analysis on the combined endpoint (Section 5.2.2) uses a summary endpoint, but which may not always be interpretable. It also renders impossible assessment of the association between the endpoints. As a third and appealing alternative, a joint analysis models both endpoints simultaneously and accommodates association between them. Conveniently, a test for the overall treatment effect based on both endpoints can be done using likelihood ratio tests.

The two preferred models in the univariate analyses, Weibull-normal for tail intensity and Weibull-gamma-normal for tail length are now combined into a joint model by assuming that the normal random effects are correlated. Table 5.4 presents the results. The contrasts of interest based on each endpoints separately as well as on the overall effect based on both endpoints is provided. The estimates are slightly different from the univariate analyses. The three contrasts (low *versus* medium, low *versus* high, and medium *versus* high) have p-values of (0.5242,0.0556,0.1959) and (0.0055,0.0001,0.0119) based on the first and second endpoints and (0.0302,0.0001,0.0302) based on the two endpoints combined. The correlation between the two random intercepts was highly significant, and estimated as 0.6049 (s.e. 0.098). The intraclass correlations were estimated as 0.1991 (s.e. 0.02704) and 0.04180 (s.e. 0.0089) for tail length and tail intensity, respectively. As a result, the pairwise correlation is estimated as 0.05499 (s.e. 0.0129).

5.3 Simulation Study

A set of simulations was conducted to evaluate the performance of the different models in terms of the type I and II error rates, as well as bias of the parameter estimates. Two batches of simulation were done for two levels of overdispersion. We considered two treatment groups: active and vehicle control. In the fist set of simulations, we assume there are 3 animals, hence 9 slides, and a total of 450 cells in each treatment group. In contrast, in the second batch, 6 animals and 18 slides each with 10 cells, are considered. Two responses were generated, one with and the other without overdispersion. The first response $Y_{1ij} \sim \text{Weibull}(\rho_1, \theta_{ij} e^{\eta_{1ij}})$ follows a Weibull-gamma-normal,

Table 5.4: *Comet Assay Study. Joint Model, Weibull-normal model for tail intensity and Weibull-gamma-normal for tail length*

		Tail Length		Tail Intensity		Overall	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value	G^2	p-value
Veh.	β_0	-29.0574(0.6537)	0.0001	-2.4620(0.0763)	0.0001	–	–
Low *vs.* veh.	β_1	-11.1543(0.4083)	0.0001	-2.8064(0.0893)	0.0001	248	0.0001
Med.*vs.* veh.	β_2	-11.3575(0.4107)	0.0001	-3.0502(0.0902)	0.0001	256	0.0001
High *vs.* veh.	β_3	-11.7724(0.4154)	0.0001	-3.2701(0.0912)	0.0001	264	0.0001
Pos. C.*vs.* veh.	β_4	-8.9600(0.4400)	0.0001	-1.7826(0.1056)	0.0001	171	0.0001
Low *vs.* Med.	β_5	-0.2032(0.3177)	0.5242	-0.2438(0.0854)	0.0055	7	0.0302
Low *vs.* High	β_6	-0.6181(0.3181)	0.0556	-0.4637(0.0856)	0.0001	25	0.0001
Med. *vs.* High	β_7	-0.4149(0.3181)	0.1959	-0.2199(0.0854)	0.0119	7	0.0302
log of Weib.P	ρ	10.0336(0.2210)	0.0001	1.4152(0.0189)	0.0001	–	–
s.d. of RE	$\sqrt{d}$	0.9227(0.0797)	0.0001	0.2133(0.0239)	0.0001	–	–
log of OD par.	α	1.0052(0.0517)	0.0001	–	–	–	–
Correlation	r	0.6049(0.0979)			0.0001		
-2 loglik				61824			

while $Y_{2ij} \sim \text{Weibull}(\rho_2, e^{\eta_{2ij}})$ follows a Weibull-normal, with further

$$\begin{aligned} \eta_{kij} &= \beta_{k0} + \beta_{k1} T_{ij} + b_{ki}, \qquad (k = 1, 2) \\ \theta_{ij} &\sim \text{Gamma}\left(\alpha, \frac{1}{\alpha}\right). \end{aligned}$$

T_{ij} is the indicator for the treatment group. Random effects are correlated and follow (5.1). Different correlation levels between the random effects as well as different overdispersion level were considered to gauge the impact of these characteristics. We set the random-effects standard deviations to $d_1 = d_2 = 0.2$ and the Weibull shape parameters to $\rho_1 = \rho_2 = 0.4$. The correlation r ranges over 0.9, 0.6, and 0.3 in both sets of simulations. Because interest lies in assessing the type I and II error rates, the data are generated under the null ($\beta_{10} = \beta_{11} = -1$ and $\beta_{21} = \beta_{21} = -1$) for the type I error rate, and under the alternative ($\beta_{10} = -1$, $\beta_{11} = -1.3$, $\beta_{20} = -1$, and $\beta_{21} = -1.3$) to assess the type II error rate. An overdispersion level of $\alpha = 0.8$ is used for the first set of simulation and $\alpha = 1.5$ for the second one. A total of 200 such datasets is generated per run. The two responses are analyzed separately using: (1) a traditional model, i.e., analysis of variance on the summary measure (mean) of the log-transformed response; (2) a classical Weibull model; (3) a Weibull-normal model;

(4) a Weibull-gamma-normal model; and finally (5) a joint model.

Simulation Results

The first simulation run is summarized in Table 5.5–5.7, with the rest deferred to the Appendix B.3. Generally, the type I error rate for all models was approximately the nominal one, except for the classical Weibull model. This could be ascribed to the independence assumption between the outcomes in this model. Indeed, ignoring the correlation may underestimate the standard errors (see Table 5.5). This has an adverse impact on the assessment of treatment effect, in the sense that a compound can easily erroneously be declared toxic. The error rate is higher for the first response where the hierarchical structure and overdispersion are omitted, in contrast to the second response.

We now turn to the power of the test. Analyzing the two responses using the various appropriate models has higher power when compared to the traditional model. The discrepancy between the proper and traditional models increases with decreasing variance of the random effects (results not presented here). This is not surprising because, when the variability between clusters is high, then the measurement within a cluster are similar. In that case, summarizing the observations has little impact. The shape parameter has an impact as well. When it gets smaller, the density becomes more skewed and the traditional approach, relying on normality and hence symmetry, drifts apart. On the other hand, the underestimated standard error when the simple classical model was employed not only inflated the type I error rate, it also exaggerates the power of the test.

The parameter estimates are also biased and the bias was higher for the estimates of the first response with overdispersion. When the first response is analyzed with the Weibull-normal model that does not account for overdispersion, the power of the test was lower and the parameter estimates still biased, though they were slightly better than under the classical Weibull model. In fact, the power of the test for the traditional model was even better. On the other hand, analyzing the data coming from the Weibull-normal model by using the Weibull-gamma-normal, leads to the same results in terms of the type I error rate, the power of the test as well as the parameter estimates, underscoring the importance of accommodating overdispersion. Given the elaborate nature of the joint analysis, it is not surprising that some convergence problems emerge. Its power is higher than that from the univariate analyses. The rise in power increases with decreasing correlation between the random effects. This

Table 5.5: *Simulation result 1.1. Type I error rates, power of the tests and parameter estimates in (1) Analysis of variance (Trad.), (2) the Weibull model (W), (3) the Weibull-normal model, (4) the Weibull-gamma-normal model (WGN), (5) the Joint model (JWGN). Correlation between* $r = 0.3$ *and* $\alpha = 0.8$.

	Resp.	Par.	Trad.	W	WN	WGN	JWGN
Power	1		0.38	0.57	0.26	0.435	0.448
	2		0.565	0.895	0.655	0.655	0.642
	Comb.						0.743
Est.(s.e.)	1	β_{10}		-0.9635(0.0575)	-0.9977(0.0901)	-1.0049(0.1129)	-1.006(0.1171)
		β_{11}		-1.1066(0.0596)	-1.1443(0.09171)	-1.2977(0.1133)	-1.3073(0.1176)
		$\beta_{11} - \beta_{10}$		-0.1430(0.0671)	-0.1466(0.1173)	-0.2929(0.1528)	-0.3000(0.1586)
	2	β_{20}		-0.9965(0.0598)	-1.0085(0.0854)	-1.0085(0.0854)	-1.0079(0.0865)
		β_{21}		-1.2765(0.0646)	-1.2934(0.0890)	-1.2934(0.0890)	-1.3007(0.0892)
		$\beta_{21} - \beta_{20}$		-0.2801(0.0671)	-0.2849(0.1087)	-0.2849(0.1087)	-0.2918(0.1087)
Type I	1		0.05	0.335	0.065	0.06	0.065
	2		0.045	0.26	0.05	0.05	0.0365
	Comb.d						0.048
Est.(s.e.)	1	β_{10}		-0.9580(0.0574)	-0.9917(0.0894)	-0.9929(0.1116)	-0.9985(0.1145)
		β_{11}		-0.9728(0.0578)	-1.0050(0.0897)	-0.9941(0.1116)	-0.9911(0.1140)
		$\beta_{11} - \beta_{10}$		-0.0148(0.0670)	-0.0134(0.1162)	-0.0012(0.1499)	0.0101(0.1565)
	2	β_{20}		-0.9783(0.0595)	-0.9911(0.0859)	-0.9911(0.0859)	-0.9901(0.0856)
		β_{21}		-0.9927(0.0598)	-1.0039(0.0861)	-1.0039(0.0861)	-1.0053(0.0869)
		$\beta_{21} - \beta_{20}$		-0.0143(0.0667)	-0.0128(0.1096)	-0.0128(0.1096)	-0.0148(0.1120)

is to be expected because lower correlation implies that a pair of outcomes is more informative. Finally, we also noted that as the cluster size grows larger, the power of the test is higher for all models (details not given).

Table 5.6: *Simulation result 1.2. Type I error rates, power of the tests and parameter estimates in (1) Analysis of variance (Trad.), (2) the Weibull model (W), (3) the Weibull-normal model, (4) the Weibull-gamma-normal model (WGN), (5) the Joint model (JWGN). Correlation between* $r = 0.6$ *and* $\alpha = 0.8$.

	Resp.	Par.	Trad.	W	WN	WGN	JWGN
Power	1		0.465	0.58	0.25	0.48	0.424
	2		0.555	0.915	0.69	0.69	0.633
	Comb.						0.696
Est.(s.e.)	1	β_{10}		-0.9595(0.0574)	-0.9940(0.0901)	-0.9978(0.1116)	-1.002(0.1136)
		β_{11}		-1.1080(0.0596)	-1.1462(0.0917)	-1.3035(0.1120)	-1.2979(0.1132)
		$\beta_{11} - \beta_{10}$		-0.1485(0.0671)	-0.1522(0.1173)	-0.3057(0.1510)	-0.2945(0.1533)
	2	β_{20}		-0.9903(0.05969)	-1.0028(0.0835)	-1.0028(0.0835)	-1.011(0.0838)
		β_{21}		-1.2823(0.0647)	-1.2970(0.0874)	-1.2970(0.0874)	-1.2967(0.0885)
		$\beta_{21} - \beta_{20}$		-0.2920(0.0672)	-0.2942(0.1060)	-0.2942(0.1060)	-0.2811(0.1073)
Type I	1		0.04	0.295	0.05	0.07	0.0666
	2		0.045	0.265	0.07	0.07	0.0588
	Comb.						0.0504
Est.(s.e.)	1	β_{10}		-0.9594(0.0574)	-0.9933(0.0898)	-0.9950(0.1106)	-0.9940(0.1122)
		β_{11}		-0.9775(0.0578)	-1.0114(0.0901)	-1.0075(0.1105)	-1.0020(0.1110)
		$\beta_{11} - \beta_{10}$		-0.0180(0.0670)	-0.0181(0.1168)	-0.0125(0.1485)	-0.0073(0.1537)
	2	β_{20}		-0.9855(0.0596)	-0.9980(0.0856)	-0.9980(0.0856)	-0.9980(0.0848)
		β_{21}		-0.9920(0.0598)	-1.0037(0.0858)	-1.0037(0.0858)	-0.9990(0.0851)
		$\beta_{21} - \beta_{20}$		-0.0065(0.0667)	-0.0057(0.1090)	-0.0065(0.1089)	-0.0005(0.1080)

Table 5.7: *Simulation result 1.3. Type I error rates, power of the tests and parameter estimates in (1) Analysis of variance (Trad.), (2) the Weibull model (W), (3) the Weibull-normal model, (4) the Weibull-gamma-normal model (WGN), (5) the Joint model (JWGN). Correlation between* $r = 0.9$ *and* $\alpha = 0.8$.

	Resp.	Par.	Trad.	W	WN	WGN	JWGN
Power	1		0.445	0.595	0.255	0.505	0.4605
	2		0.565	0.915	0.675	0.675	0.6619
	Comb.						0.6056
Est.(s.e.)	1	β_{10}		-0.9616(0.0574)	-0.9959(0.0897)	-1.0016(0.1126)	-1.011(0.1185)
		β_{11}		-1.1117(0.0596)	-1.1486(0.0913)	-1.3082(0.1131)	-1.302(0.1171)
		$\beta_{11} - \beta_{10}$		-0.1501(0.0671)	-0.1527(0.1166)	-0.3066(0.1525)	-0.2997(0.1600)
	2	β_{20}		-0.9902(0.0597)	-1.0029(0.0861)	-1.0029(0.0861)	-1.0105(0.0883)
		β_{21}		-1.2838(0.0647)	-1.3016(0.0899)	-1.3016(0.0899)	-1.2990(0.0905)
		$\beta_{21} - \beta_{20}$		-0.2936(0.0672)	-0.2988(0.1099)	-0.2988(0.1099)	-0.2860(0.1125)
Type I	1		0.08	0.33	0.08	0.085	0.0909
	2		0.06	0.245	0.08	0.08	0.0666
	Comb.						0.0512
Est.(s.e.)	1	β_{10}		-0.9593(0.0574)	-0.9936(0.0896)	-0.997(0.1117)	-1.0019(0.1095)
		β_{11}		-0.977(0.0579)	-1.0101(0.0899)	-1.005(0.1116)	-1.0064(0.1083)
		$\beta_{11} - \beta_{10}$		-0.0180(0.0670)	-0.0164(0.1166)	-0.0081(0.1501)	-0.0083(0.1555)
	2	β_{20}		-0.9867(0.05965)	-0.99948(0.0859)	-0.9995(0.0859)	-1.013(0.0853)
		β_{21}		-0.9889(0.05976)	-1.0011(0.08598)	-1.0011(0.08598)	-0.9902(0.0825)
		$\beta_{21} - \beta_{20}$		-0.00214(0.0667)	-0.00164(0.1094)	-0.00165(0.1094)	0.0139(0.1093)

5.4 Concluding Remarks

Co-primary endpoints are commonly used to assess the toxic effect of a certain compound in toxicological studies. Univariate analyses are often done on each endpoint separately; but this leads to multiple inferences. Joint modeling of the endpoints is appealing to make overall inferences as well as to capture the association among the outcomes. In this chapter, joint model using a random-effect was presented in a bivariate setting with hierarchically clustered and overdispersed non-Gaussian continuous outcomes. Thus, the model accounts for: (1) overdispersion; (2) repeated measures over time; (3) and the multivariate nature of the outcomes.

Two Weibull-gamma-normal models were combined using bivariate normally distributed random effects. This is a simple and relatively less restrictive approach compared to a shared parameter model and it can be easily implemented in standard software like in the SAS procedure NLMIXED.

It was applied to the comet assay data which exhibit two outcomes namely tail length and tail intensity. Univariate analyses indicate that a model with overdisperion (Weibull-gamma-normal) is necessary for tail length and a model without overdispersion (Weibull-normal) is sufficient for the tail intensity. The contrast between low, medium, and high dose level using the two endpoints leads to different conclusion.

Chapter 6

Finite-mixture and Zero-inflated Models for Hierarchically Clustered and Overdispersed Outcomes

As already explained in Chapter 3, due to the prescribed mean-variance relationship, overdispersion is a commonly encountered phenomenon in non-Gaussian data. It can be driven by different factors: multimodality in the data due to the presence of subpopulations within an overall population, a lack of information about which subpopulation an observation belongs to, highly skewed nature of the data, which could be partly due to sets of outlying observations, the presence of excess zeros in the data, and many more.

In general in statistical modeling, because of their usefulness as a flexible method of modeling, finite mixture models received increasing attention over the years, both from a practical and a theoretical point of view. It occupies an interesting niche between parametric and non-parametric approaches to statistical estimation. As explained by Jordan and Xu (1995) mixture-model-based estimation approaches are parametric in that parametric forms are specified for the component density functions, but they can also be regarded as non-parametric by allowing the number of components K to grow. Hence, the mixture models have much of the flexibility of

non-parametric models, while retaining some of the advantages of parametric approaches, such as keeping the dimension of the parameter space down to a reasonable size. They have been useful in modeling quite complex distributions through an appropriate choice of its components to accurately represent the local areas of support of the true distribution. It can thus handle situations where a single parametric family is unable to provide a satisfactory model for local variation in the observed data.

In Chapters 4 and 5, in line with Molenberghs *et al* (2010), overdispersion is accounted for through a continuous conjugate random effect. However, the assumed conjugate distribution can be misspecified. In these circumstances, mixture models may be useful. Note that in the combined model, the overdispersion is accounted for through a random effect and this random effect is typically assumed to have a continuous, conjugate distribution. This can be regarded as a mixture of models where each observation is allowed to have its own model, with the number of mixture components equal to the number of observations. In this chapter, we will consider three situations in which a finite mixture model is useful. First, the overdispersion random effect can have a complex distribution, different from the previously assumed conjugate distribution. In this case, assuming a mixture of conjugate distributions for the random effect can give further flexibility to the combined model. Second, the source of the overdispersion could be due to the existence of unobserved subpopulations (latent classes). In this case, the overdispersion random effect can be well expressed by some discrete distribution, instead of a continuous distributed random effect. Third, the source may also stem from excess zeros, which can be accounted for by zero-inflated models; these in turn may be regarded as mixtures of models. The different situations will be investigated for the comet assay data. These results are summarized in Ghebretinsae *et al* (2013).

Section 6.1 is devoted to gamma mixtures, while Section 6.2 focuses on Weibull mixtures. Zero inflation is the topic of Section 6.3. An application to the comet data is the subject of Section 6.4.

6.1 A Mixture of Gamma Distributions for the Overdispersion Random Effect

One way to account for overdispersion is through the use of random effects. In Section 3.1.4, a general model family proposed by Molenberghs *et al* (2010) for modeling overdispersed and correlated data is presented. A model with a normal random effect

to handle the hierarchy in the data and a conjugate random effect to account for additional overdisperion in the response is used. In the context of comet data, since the outcomes are skewed, non-negative and continuous, which is similar to time-to-event data, a Weibull-type version of the combined model is used.

Now, while the gamma frailty (random effect) is convenient as a specification for the overdispersion because of the conjugacy of the gamma and Weibull distribution, misspecification of this distribution is possible. Alternatives to the gamma distribution can be considered but are computationally more complex, because of the lack of the conjugacy property. In this section, a mixture of conjugate distributions is considered for the specification of the overdispersion term, as an alternative to the previously proposed gamma frailty. In this way, flexibility is added to the overdispersion distribution, and misspecification of the overdispersion can be checked. Deviations from a single conjugate distributions are allowed for by the use a mixture distribution, and the property of conjugacy can still be employed to ease computations. Let us first consider the situation where clustering is not taken into account. In this case, model (3.13)–(3.16) simplifies to the gamma frailty (Weibull-gamma) model, given by:

$$f(y_i|\theta_i) = \lambda\rho\theta_i y_i^{\rho-1} e^{\boldsymbol{x}_i'\boldsymbol{\xi}} e^{-\lambda y_i^{\rho}\theta_i e^{\boldsymbol{x}_i'\boldsymbol{\xi}}}, \tag{6.1}$$

$$f(\theta_i|\alpha) = \frac{1}{\left(\frac{1}{\alpha}\right)^{\alpha}\Gamma(\alpha)}\theta_i^{\alpha-1}e^{-\alpha\theta_i}. \tag{6.2}$$

This model is now extended by replacing (6.2) by a mixture of gamma distributions. It is produced by combining two or more gamma distributions using mixing parameters that represent the proportion of mixing of the components. Let us assume the distribution of the random effect (θ_i) is a mixture of two gamma distributions instead of one. In this case, (6.2) is replaced by:

$$f(\theta_i|\alpha_1,\alpha_2) = P.f(\theta_i|\alpha_1) + (1-P).f(\theta_i|\alpha_2),$$

where P is the proportion of mixing of the components. The marginal distribution can easily be derived by integrating over both random effects, as it is given by:

$$\begin{aligned} f(y_i) &= \int f(y_i|\theta_i)f(\theta_i|\alpha_1,\alpha_2)d\theta_i, \\ &= P\int f(y_i|\theta_i)f(\theta_i|\alpha_1)d\theta_i + (1-P)\int f(y_i|\theta_i)f(\theta_i|\alpha_2)d\theta_i. \end{aligned}$$

This can be extended to account for clustering by incorporating a normal random effect in the linear predictor $\boldsymbol{x}'_{ij}\boldsymbol{\xi}$, that is, $\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$.

6.2 A Discrete Mixture of Weibull Distributions

Misspecification of the conjugate random effect was addressed in Section 6.1 through gamma mixtures. On the other hand, it is also possible to explain overdispersion by considering only a few discrete values instead of infinitely many values, as represented by a continuous random effect. That is, the overdispersion may be derived by some subpopulations or latent classes which lead to the mixture models. Here, mixtures based on the Weibull and Weibull-normal models are considered and the connection with the combined model is outlined.

Consider a Weibull-type combined model and suppose the overdispersion random effect θ_{ij} has only two discrete values, θ_1 or θ_2.

$$\begin{aligned} f(y_{ij}|\theta_1) &= \text{Weibull}(\rho, \lambda\theta_1 \exp(\eta_{ij})) \\ f(y_{ij}|\theta_2) &= \text{Weibull}(\rho, \lambda\theta_2 \exp(\eta_{ij})) \end{aligned}$$

Now, these values can be absorbed into the scale parameter, $\lambda \exp(\eta_{ij})$. This implies that the scale parameter of the Weibull distribution is either $\lambda_1 \exp(\eta_{ij})$ or $\lambda_2 \exp(\eta_{ij})$. So, previous expression can be rewritten as:

$$\begin{aligned} f(y_{ij}|\lambda_1) &= \text{Weibull}(\rho, \lambda_1 \exp(\eta_{ij})) \\ f(y_{ij}|\lambda_2) &= \text{Weibull}(\rho, \lambda_2 \exp(\eta_{ij})) \end{aligned}$$

It leads to a mixture of two Weibull-normal models with different scale parameters but the same shape parameter ρ. A randomly picked observation either belongs to population 1, a Weibull$(\rho, \lambda_1 \exp(\eta_{ij}))$ or population 2, a Weibull$(\rho, \lambda_2 \exp(\eta_{ij}))$:

$$f(y_{ij}) = \begin{cases} \text{Weibull}(\rho, \lambda_1 \exp(\eta_{ij})) & \text{with probability} \quad P, \\ \text{Weibull}(\rho, \lambda_2 \exp(\eta_{ij})) & \text{with probability} \quad 1-P, \end{cases}$$

where P is the proportion of the first component and $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$. This is a special case of the combined model, taking a discrete distribution for the overdispersion random effect. This can further be extended by allowing the shape parameter to vary, leading to a full mixture model.

In general, mixtures of Weibull distributions are produced by combining two or more Weibull distributions using mixing parameters that represent the proportion of mixing of the components:

$$f(y_{ij}) = \sum_{l=1}^{K} P_l \lambda_l \rho_l y_{ij}^{\rho_l - 1} e^{\eta_{ij}} e^{-\lambda_l y_{ij}^{\rho_l} e^{\eta_{ij}}},$$

where $P_l > 0$ is the proportion of the mixing of the l^{th} component, K is the number of components and

$$\sum_{l=1}^{k} P_l = 1.$$

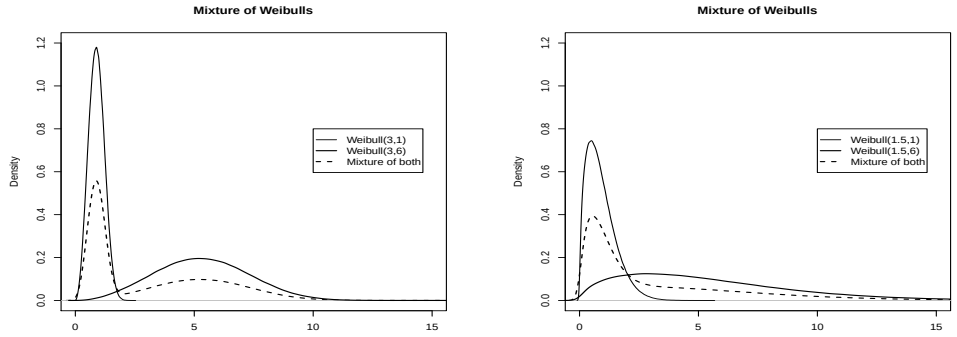
The different components may correspond to meaningful subpopulations. For example, they may correspond to the different species of the rat, or may not be identified but rather introduced to allow for greater flexibility in modeling a heterogenous population that is apparently unable to be modeled as a single component distribution.

In modeling finite mixture models, often mixtures of normal densities are employed as any continuous distribution can be approximated well by a finite mixture of normal densities. However, mixtures of other distributions can also be used. The Weibull distribution by itself is flexible with a variety of shapes and using mixtures of Weibull distributions adds more flexibility. Figure 6.1 present different densities resulting from mixing different pairs of Weibull distributions with equal proportions.

6.3 Zero-inflated Mixture Models

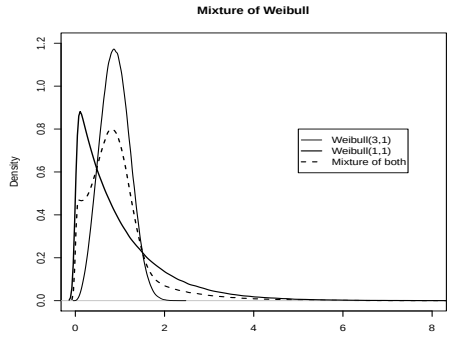
When the data have a pronounced excess of zeros, none of the different Weibull models considered so far sufficiently account for these zeros. Further extension is needed, motivating the zero-inflated models.

Zero-inflated models are commonly used to model counts data with an excess of zeros. The idea in the zero-inflated Poisson model (ZIP) is that it assumes outcomes to emanate from two processes. One process models zero inflation by including a proportion of extra zeros and another proportion of zeros coming from the regular Poisson distribution. However, the classical Weibull model accounts for values $Y_{ij} > 0$. To account for the possible presence of zeros, one component is added to the distribution, a point mass at zero. A zero-inflated Weibull model (ZIW) has, like the

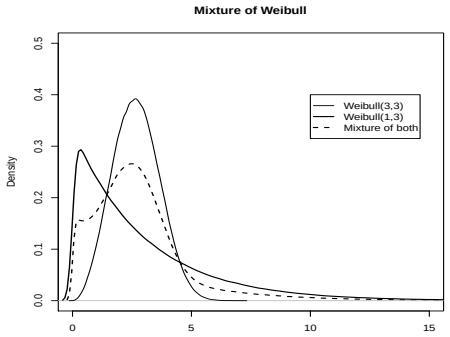


(a)

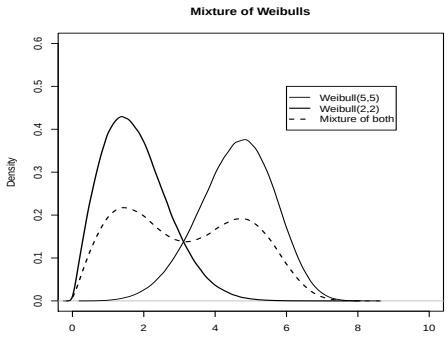
(b)



(c)



(d)



(e)

Figure 6.1: *Mixtures of Weibull densities with different shape and scale parameters.*

zero-inflated Poisson model, two compartments. But the zeros in the ZIP may come from the point mass and the Poisson component, which is not the case for the zero-inflated Weibull model, since the zeros are entirely from the point mass. Similarly, the zero-inflated Weibull-normal (ZIWN) and zero-inflated combined model or zero-inflated Weibull-gamma-normal model (ZIWGN) are produced by incorporating an additional component of a point mass at zero to the Weibull-type GLMM (Weibull-normal) and combined model (Weibull-gamma-normal). Let us consider the zero-inflated combined model. The density function is given by:

$$f(y_{ij}|\theta_{ij}, \boldsymbol{b_i}) = \begin{cases} \phi(\eta_{1ij}) & \text{if} \quad y_{ij} = 0, \\ \lambda\rho\theta_{ij} y_{ij}^{\rho-1} e^{\eta_{2ij}} e^{-\lambda y_{ij}^{\rho}\theta_{ij} e^{\eta_{2ij}}} & \text{if} \quad y_{ij} > 0, \end{cases}$$

where $\eta_{1ij} = \boldsymbol{x}_{ij}\boldsymbol{\xi}_1$ and $\eta_{2ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi}_2 + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$. The additional component for the zero is also allowed to be a function $[\phi(\cdot)]$ of covariates. In this case, it will be dose group. If the proportion of zeros across the dose levels are informative by themselves, inference may be drawn based on the combined effects of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. If the overdispersion random effect (θ_{ij}) is dropped, it reduces to a zero-inflated Weibull-normal model and if the hierarchical random effect ($\boldsymbol{b_i}$) is further dropped, it reduces to a zero-inflated Weibull model. Similarly, a zero-inflated mixture of Weibulls and zero-inflated mixture of Weibull-normal models extends the mixture of Weibull distributions of Section 6.2. They have three compartments: a point mass at zero and two Weibull models or Weibull-normal models. The zero-inflated Weibull-normal model is given by:

$$f(y_{ij}|\boldsymbol{b_i}) = \begin{cases} \phi(\eta_{1ij}) & \text{if} \quad y_{ij} = 0, & \\ \lambda_1\rho y_{ij}^{\rho-1} e^{\eta_{2ij}} e^{-\lambda_1 y_{ij}^{\rho} e^{\eta_{2ij}}} & \text{if} \quad y_{ij} > 0 & \text{with prob.} \quad \text{P}, \\ \lambda_2\rho y_{ij}^{\rho-1} e^{\eta_{2ij}} e^{-\lambda_2 y_{ij}^{\rho} e^{\eta_{2ij}}} & \text{if} \quad y_{ij} > 0 & \text{with prob.} \quad \text{1-P}. \end{cases}$$

6.4 Application to the Comet Data

Also here, the primary aim is to assess the toxicity of 1,2-Dimethylhydrazine dihydrochloride at different dose levels. The two sets of comet assay data presented in Section 2.1 are analyzed using the statistical models of Sections 6.1–6.3. The mixture of Weibull distributions and the mixture of Gamma distributions are applied to the first set, which apparently does not have zeros, likewise, zero-inflated models are applied to the second set with excess zeros. All analyses are done on tail length and

the linear predictor part of the models is given by (4.7). In what follows, the results are presented.

6.4.1 Mixture of Gamma Distributions

To add flexibility to the distributional assumption for the overdispersion, a Weibull-type combined model with a mixture of two conjugate gamma distributions for overdispersion is applied to the comet data. Some of the results for models with and without a normal hierarchical random effects are presented in Table 6.1. The Weibull-gamma model is presented, together with the Weibull-mixture of gamma in the top panel. Extensions of these models, by the inclusion of a normal random effect to account for hierarchy in the data, is given in the lower panel. Using a mixture of conjugate distributions has improved the fit slightly in the two sets of models. It indicates that almost all of the observation-specific random effects come from one component and only a few from the other component. The other note that needs to be made is the identifiability of the intercept and scale parameter in modeling the combined model. The Weibull-gamma-normal model in Tables 6.1 and 6.2 are the same, except that the scale parameter is fixed to one. The only difference between the two models is in the estimate of the intercept, $\exp(-30.93) = 0.005 \cdot \exp(-25.63)$ and the rest of the parameter estimates remain the same. This indicates that the intercept and scale parameter are only jointly identifiable. At the same time, it does not have an effect on the inference drawn, as the comparison of interest remains unaffected.

6.4.2 Mixture of Weibull Models

A mixture of Weibull models was employed to assess if the overdispersion could be explained by a few unobserved components. In this case, a mixture of two Weibull models was considered. Analyses with and without a normal random effect to account for the clustering are done. A summary of the result is presented in Table 6.2. The upper panel corresponds to the Weibull model, Weibull-mixture model with two mixture components, and the Weibull-gamma model, which can be seen as a mixture of with infinite number of components. The lower panel corresponds with these three models, extended by a normal random effect to take into account the clustering as well. Comparing the models with and without the normal hierarchical random effect, it can be seen that the models accounting for the hierarchical structure have a better fit, underscoring the importance of the hierarchical structure. In both sets of models, we clearly see the discrete mixture model improved the fit in terms of the likelihood.

Table 6.1: *Comet Assay Study. Parameter estimates, standard error, and significance level for the regression coefficient in (1) the Weibull-gamma model, Model 1, (2) the Weibull mixture of gamma model, Model 2, (3) the Weibull-gamma-normal model, Model 3, (4) the Weibull-mixture of gamma-normal model, Model 4.*

		Model 1		Model 2	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value
Veh.	β_0	-27.2624(0.6211)	0.0001	-22.3976(1.7424)	0.0001
Low *vs.* Med.	$\beta_2 - \beta_1$	-0.1862(0.08511)	0.0288	-0.1893(0.08670)	0.0290
Low *vs.* High	$\beta_3 - \beta_1$	-0.5556(0.08621)	0.0001	-0.5646(0.08798)	0.0001
Med. *vs.* High	$\beta_3 - \beta_2$	-0.3694(0.08699)	0.0001	-0.3753(0.08889)	0.0001
Weib.P	ρ	9.4499(0.2242)	0.0001	9.6691(0.2426)	0.0001
OD par.1	α_1	0.8619(0.04527)	0.0001	0.00012(0.00008)	0.1474
OD par.2	α_2	–	–	0.8286(0.0446)	0.0001
prop.	p	–	–	0.9958(0.00722)	0.0001
-2loglik.		28904		28891	
		Model 3		Model 4	
Effect	Par.	Estimate(s.e.)	p-value	Estimate(s.e.)	p-value
Veh.	β_0	-30.9295(0.7264)	0.0001	-24.4192(1.6681)	0.0001
Low *vs.* Med.	$\beta_2 - \beta_1$	-0.2174(0.3398)	0.5241	-0.2204(0.3557)	0.5373
Low *vs.* High	$\beta_3 - \beta_1$	-0.6648(0.3403)	0.0543	-0.6897(0.3564)	0.0565
Med. *vs.* High	$\beta_3 - \beta_2$	-0.4474(0.3402)	0.1923	-0.4693(0.3563)	0.1916
Weib.P	ρ	10.7072(0.2474)	0.0001	11.1825(0.2829)	0.0001
s.d. of RE	$\sqrt{d}$	0.9881(0.0859)	0.0001	1.0354(0.0906)	0.0001
OD par.1	α_1	0.8932(0.0463)	0.0001	0.8313(0.0446)	0.0001
OD par.2	α_2	–	–	0.0002(0.0001)	0.0001
prop.	p	–	–	0.9996(0.0006)	0.0001
-2loglik.	–	28069		28037	

It also has some impact on the estimate as well as the precision and this in turn had impact on the conclusion inferred for some contrasts of interest. For example, in the contrast between low dose and high dose the significance level changed from p=0.1357 for the Weibull-normal model to p=0.0395 for the mixture model. Similarly, the comparison of low versus medium changed from p=0.0431 for the Weibull model to p=0.001 for the mixture of Weibull models. Based on the mixture model, the majority of the observations are allegedly from one component and a small amount of about 1.8 % is from the second component. The observed and predicted densities are given in Figure 6.2. The overdispesion seems to be driven by some small set

of large/extreme observations. When the overdispersion is accounted for through a gamma random effect, the models have still a better performance, showing that the overdispersion may not be sufficiently explained by just two components. In modeling the mixture model, further allowing the shape parameter to vary did not improve the fit.

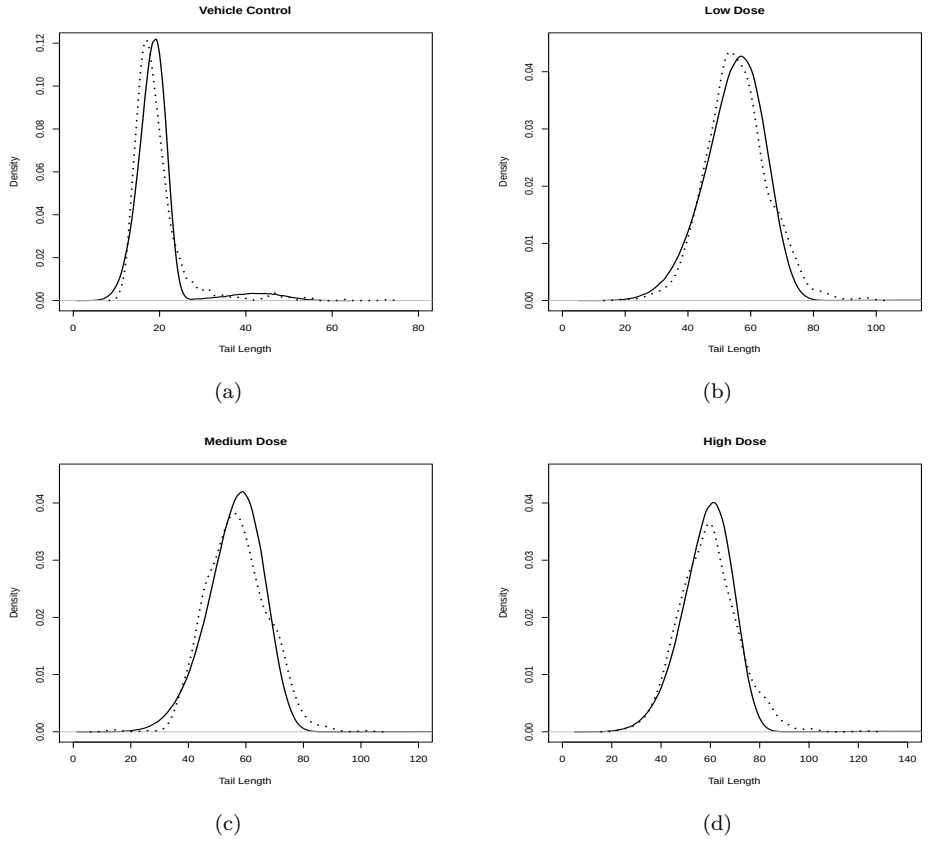


Figure 6.2: *Observed (dotted line) and predicted estimate (solid line), from the Mixture of Weibull-normal models, for tail length by dose groups.*

6.4.3 Zero-inflated Models

In this analysis, the different zero-inflated models, as proposed in Section 6.3, were applied on the second set of comet data, that has a non-negligible amount of zeros

Table 6.2: *Comet Assay Study: tail length. Parameter estimates, standard error, and significance level for the regression coefficient in (1) the Weibull model, Model 1, (2) the mixture of two Weibull models, Model 2, (3) the Weibull-gamma model, Model 3, (4) the Weibull-normal model, Model 4, (5) the mixture of Weibull-normal models, Model 5, (6) the Weibull-gamma-normal model, Model 6.*

		Model 1		Model 2		Model 3	
Effect	Par.	Est.(s.e.)	p-val.	Est.(s.e.)	p-val.	Est.(s.e.)	p-val.
Veh.	β_0	-17.93(0.03)	0.0001	-21.73(0.30)	0.0001	-25.93(0.08)	0.0001
Low *vs.* Veh.	β_1	-3.55(0.05)	0.0001	-6.35(0.10)	0.0001	-10.54(0.27)	0.0001
Med. *vs.* Veh.	β_2	-3.64(0.05)	0.0001	-6.52(0.10)	0.0001	-10.72(0.27)	0.0001
High *vs.* Veh.	β_3	-3.85(0.05)	0.0001	-6.78(0.10)	0.0001	-11.09(0.28)	0.0001
Pos.C. *vs.* Veh.	β_4	-2.70(0.06)	0.0001	-5.17(0.09)	0.0001	-8.51(0.23)	0.0001
Low *vs.* Med.	β_5	-0.10(0.05)	0.0431	-0.16(0.05)	0.0010	-0.19(0.09)	0.0288
Low *vs.* High	β_6	-0.30(0.05)	0.0001	-0.43(0.05)	0.0001	-0.56(0.09)	0.0001
Med. *vs.* High	β_7	-0.21(0.05)	0.0001	-0.27(0.05)	0.0001	-0.37(0.09)	0.0001
Weib.Shape	ρ	4.01(0.04)	0.0001	5.76(0.07)	0.0001	9.45(0.22)	0.0001
Weib.Scale1	λ_1	179.96(25.31)	0.0001	1	–	0.26(0.18)	0.1420
Weib.Scale2	λ_2	–	–	100.03(14.22)	0.0001	–	–
Overdis.	α	–	–	–	–	0.86(0.05)	0.0001
prop.	P	–	–	0.02(0.002)	0.0001	–	–
-2loglik.		30867		29242		28904	

		Model 4		Model 5		Model 6	
Effect	Par.	Est.(s.e.)	p-val.	Est.(s.e.)	p-val.	Est.(s.e.)	p-val.
Veh.	β_0	-23.80(0.14)	0.0001	-25.65(0.39)	0.0001	-25.63(0.23)	0.0001
Low *vs.* veh.	β_1	-4.50(0.22)	0.0001	-7.43(0.24)	0.0001	-11.94(0.44)	0.0001
Med.*vs.* veh.	β_2	-4.60(0.22)	0.0001	-7.60(0.24)	0.0001	-12.16(0.45)	0.0001
High *vs.* veh.	β_3	-4.83(0.23)	0.0001	-7.89(0.24)	0.0001	-12.60(0.45)	0.0001
Pos.C.*vs.* veh.	β_4	-3.48(0.27)	0.0001	-6.09(0.28)	0.0001	-9.64(0.48)	0.0001
Low *vs.* Med.	β_5	-0.10(0.22)	0.6393	-0.17(0.22)	0.4432	-0.22(0.34)	0.5242
Low *vs.* High	β_6	-0.33(0.22)	0.1357	-0.46(0.22)	0.0395	-0.66(0.34)	0.0542
Med. *vs.* High	β_7	-0.23(0.22)	0.3029	-0.29(0.22)	0.1897	-0.45(0.34)	0.1922
Weib.Shape	ρ	4.96(0.06)	0.0001	6.79(0.08)	0.0001	10.71(0.25)	0.0001
Weib.Scale1	λ_1	3524.37(708.51)	0.0001	1		0.005(0.004)	0.1929
Weib.Scale2	λ_2	–	–	235.94(41.67)	0.0001	–	–
s.d. of RE	$\sqrt{d}$	0.65(0.05)	0.0001	0.65(0.05)	0.0001	–	–
Overdis.	α	–	–	–	–	0.89(0.05)	0.0001
prop.	P	–	–	0.02(0.002)	0.0001	–	
-2loglik.	–	29793		28230		28069	

(10.47 %). From the exploratory analysis, the distribution of zeros varies across the dose groups: 17.47% in vehicle control, and 10, 8.67 and 6.93 % in the low, medium and high dose groups, respectively. The general trend is that the amount zeros decreases with increasing dose. In modeling, the covariate dose is therefore included in the component for zero outcomes. In the zero-inflated mixture of Weibull models, one of the scale parameter is fixed to one, for reasons of identifiability. The results with and without clustering random effect are presented in Table 6.3 (upper versus lower panel). The result for the zero component for all the zero inflated models is the same and is presented separately in Table 6.4. The estimates (percentage of zeros) can be informative by itself and could be included in the inference. In this analysis, the zero-inflated mixture of Weibulls fitted better as opposed to the zero-inflated combined model. It was consistent in both sets of models. Also here, it has an impact on the estimates and the inference drawn for some contrasts of interest. The fit of the zero-inflated mixture model is graphically illustrated in Figure 6.3.

Table 6.3: *Comet Assay Study. Parameter estimates, standard error, and significance level for the regression coefficient in (1) the zero-inflated Weibull model, Model 1, (2) the zero-inflated mixture of two Weibull models, Model 2, (3) the zero-inflated Weibull-gamma model, Model 3, (4) the zero-inflated Weibull-normal model, Model 4, (5) the zero-inflated mixture of Weibull-normal models, Model 5, (6) the zero-inflated Weibull-gamma-normal model, Model 6.*

		Model 1		Model 2		Model 3	
Effect	Par.	Est.(s.e.)	p-value	Est.(s.e.)	p-value	Est.(s.e.)	p-value
Veh.	β_0	-4.71(0.04)	0.0001	-0.59(0.08)	0.0001	-0.72(0.04)	0.0001
Low *vs.* veh.	β_1	-0.02(0.06)	0.7723	-0.13(0.07)	0.08	-0.02(0.06)	0.7856
Med.*vs.* veh.	β_2	0.02(0.06)	0.7567	-0.07(0.07)	0.3101	0.033(0.06)	0.6030
High *vs.* veh.	β_3	-0.2(0.06)	0.0004	-0.28(0.07)	0.0001	-0.22(0.06)	0.0005
Weib.Shape1	ρ_1	0.45(0.01)	0.0001	0.65(0.02)	0.0001	0.49(0.02)	0.0001
Weib.Scale1	λ_1	116.61(1.91)	0.0001	1		2.44(0.10)	0.0001
Weib.Shape2	ρ_2	–	–	0.81(0.03)	0.0001	–	–
Weib.Scale2	λ_2	–	–	17.88(1.73)	0.0001	–	–
Overdis.	α	–	–	–	–	6.71(2.5122)	0.0076
prop. of zero	P_1	0.17(0.01)	0.0001	0.17(0.01)	0.0001	0.17(0.01)	0.0001
prop. of mixture	P_2	–	–	0.60(0.02)	–	–	–
-2loglik.		8078.6		7769.3		8069.5	

		Model 4		Model 5		Model 6	
Effect	Par.	Est.(s.e.)	p-value	Est.(s.e.)	p-value	Est.(s.e.)	p-value
Veh.	β_0	-4.77(0.04)	0.0001	-0.82(0.08)	0.0001	-2.12(0.05)	0.0001
Low *vs.* veh.	β_1	-0.02(0.07)	0.7942	-0.17(0.09)	0.0839	-0.02(0.08)	0.8062
Med.*vs.* veh.	β_2	0.02(0.07)	0.8213	-0.10(0.09)	0.3027	0.03(0.08)	0.6761
High *vs.* veh.	β_3	-0.20(0.06)	0.0033	-0.32(0.09)	0.0009	-0.2(0.08)	0.0043
Low *vs.* High	$\beta_3 - \beta_1$	-0.18(0.06)	0.0060	-0.16(0.09)	0.1001	-0.21(0.08)	0.0073
Weib.Shape	ρ	0.45(0.01)	0.0001	0.66(0.03)	0.0001	0.50(0.02)	0.0001
Weib.Scale1	λ_1	124.15(2.36)	0.0001	1	0.0001	10.06(0.44)	0.0001
Weib.Scale2	λ_2	–	–	17.15(1.52)	0.0001	–	–
s.d. of RE	$\sqrt{d}$	0.09(0.03)	0.0065	0.15(0.04)	0.0003	0.12(0.04)	0.0022
Overdis.	α	–	–	–	–	5.97(2.10)	0.0062
prop. of zero	P_1	0.17(0.01)	0.0001	0.17(0.01)	0.0001	0.17(0.01)	0.0001
prop. of mixture	P_2	–	–	0.54(0.01)	0.0001	–	–
-2loglik.		8075.6		7776.0		8065.1	

Table 6.4: *Comet Assay Study. Parameter estimates, standard error, and significance level for the regression coefficient of the zero component for all the zero-inflated models in* **??** *and the observed and estimated proportion of zeros in each dose group.*

Effect	Par.	Estimate(s.e.)	P-value	Observed Prob.	Estimated Prop.
Veh.	β_1	0.4401(0.0619)	0.0001	0.17466	0.17467
Low.	β_2	0.7872(0.0554)	0.0001	0.01000	0.09999
Med	β_3	0.8566(0.0551)	0.0001	0.08666	0.08666
High	β_4	0.9543(0.0553)	0.0001	0.06933	0.06934

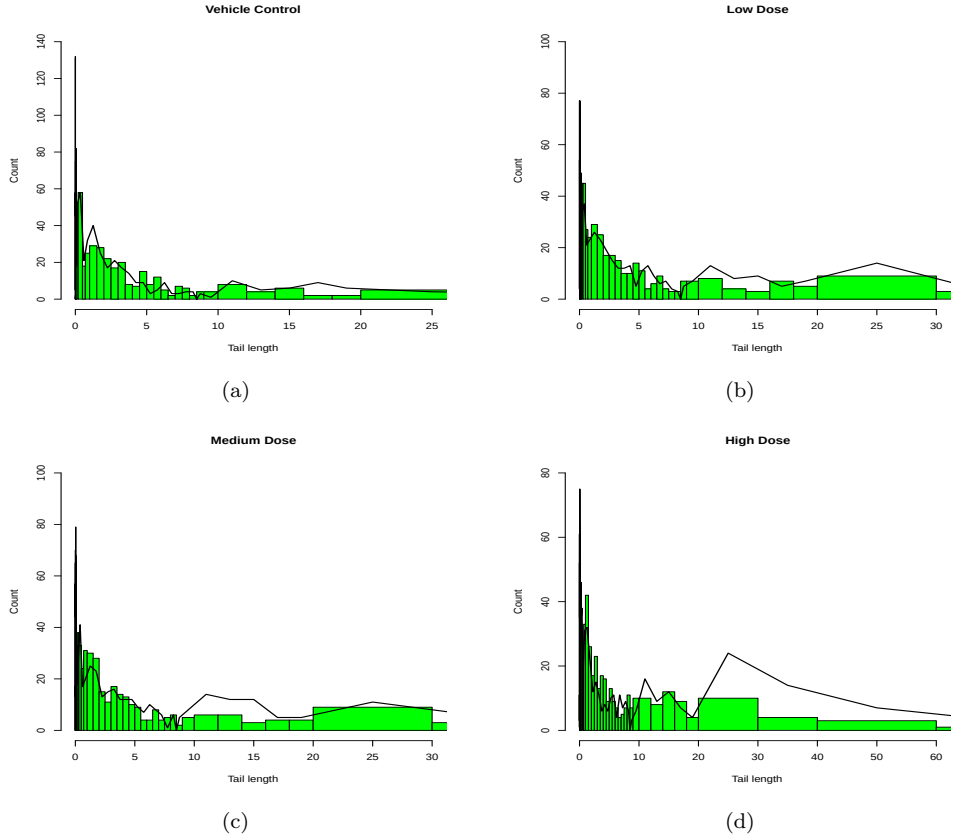


Figure 6.3: *Comet Assay Study. Observed (histograms) and predicted estimates (solid line), from the zero-inflated mixture of Weibull-normal models, for tail length by dose groups.*

6.5 Concluding Remarks

In modeling non-Gaussian data that are hierarchically structured and are overdispersed in the sense that the distributional mean-variance relationship is not fulfilled, Molenberghs *et al* (2010) proposed a general modeling framework using two random effects. In Chapters 4 and 5, a modeling framework for comet assay data was proposed which fits within this framework, but with some extension. A conjugate distribution is assumed for the overdispersion random effect. It is convenient because of conjugacy. However, misspecification of this distribution is possible. In this chapter, the use of finite mixture models is explored for comet assay data. Mixtures of conjugate distributions are considered for the overdispersion random effect to accomodate deviations from the assumed distribution. Finite mixture models are considered to asses whether the extra-variation could be explained by a small number of components. In the joint presence of excess zeros and dispersed data, a zero-inflated model was also explored.

This wide range of models was applied to the comet assay data. Here, two datasets were considered, one with and another without pronounced zeros. Finite mixture models and mixtures of the conjugate Gamma distributions were applied to the first set of the data and zero-inflated models to the other set. The use of mixture of two Gamma distributions instead of one has improved the fit. Using mixtures of Weibull-normal improved the fit but the combined model fitted better. On the other hand, the zero-inflated Weibull-normal yielded a better fit as opposed to the zero-inflated combined model for these particular datasets.

Chapter 7

Gaussian Variational Approximation for Overdispersed Generalized Linear Mixed Models

In the preceding Chapters 3–6, the proposed general framework for modeling non-Gaussian data that are hierarchically structured and overdispersed was presented. Difficulty in inference of these models is often encountered in both the Bayesian and likelihood frameworks, due to the intractable multivariate integrals in the likelihood and posterior densities. This can already be problematic for the generalized linear mixed models (GLMM), because of the integrals in the marginalized likelihood with no analytic expression that need numerical approximations. Often, this is dealt with by numerical integration using adaptive and non-adaptive Gaussian quadrature, series expansion methods including penalized quasi-likelihood and marginal quasi-likelihood, Laplace approximations, etc. Different estimation techniques have been employed when fitting the combined model. Commonly, this is done through partial marginalization in which the conjugate random effects are first integrated out, leaving the normal effects untouched. Then, one obtains the fully marginal expression by numerical integration of the normal random effect using adaptive Gaussian quadrature in standard software such as the SAS procedure NLMIXED (Molenberghs, Verbeke,

and Demétrio 2007, Molenberghs *et al* 2010). In the Bayesian framework this is done using MCMC (Ghebretinsae *et al* 2013) and pseudolikelihood estimation (Effendi, Molenberghs, and Verbeke 2010).

In this chapter, another estimation method for the combined model is proposed, providing a fast estimation method as an alternative to the existing methods. Ormerod and Wand (2012) recently introduced variational approximation in a statistical modeling framework. Variational inference has its roots in statistical physics and is used to approximating intractable mathematical expressions (Blei and Jordan 2006). The key idea is to introduce a set of approximating densities to the posteriors and to introduce them in such a way as to make their evaluation tractable. These approximations are then optimized so as to minimize the discrepancy between the approximation and the true posterior using some measure of the difference. The optimization is carried out by varying the functional parameters of these approximations, thus giving the approximation its name. While different variational approximations exist, we focus on Gaussian variational approximations, in which the conditional distribution of the random effects given the data are approximated by Gaussian distributions. Hall, Ormerod, and Wand (2011) studied the properties of Gaussian variational approximations in the setting of generalized linear mixed models.

The general idea of variational approximation is to approximate the likelihood so that the integral problem is either fully or partially solved. It is therefore basically an approximation of the integrand. When the integral problem is fully solved, it results in optimization of the resulting approximate likelihood. When the integral problem is not completely eradicated, like, for example, in a binary GLMM as will be shown later, it is still useful in reducing the dimension of the integral to one. But approximation of the integral of the new likelihood/integrand is still required, using adaptive Gaussian quadrature.

The rest of this chapter is organized as follows. The Gaussian variational approximation estimation technique is presented in Section 7.1. Its properties are investigated via three examples, i.e., an extended random-effects Weibull, Poisson, and logistic model, in Sections 7.2, 7.3 and 7.4, respectively. The Results are then discussed in Section 7.5.

7.1 Gaussian Variational Approximation Using Density Transformation (GVA)

This section presents the GVA approximation method using density transformation, as described by Ormerod and Wand (2012) and Hall, Ormerod, and Wand (2011).

7.1.1 Deriving GVA Lower Bound Likelihood

Ormerod and Wand (2012) considered the generalized linear mixed model (3.5) with the scale parameter ϕ fixed to one. The corresponding marginal log likelihood is,

$$\begin{aligned}
\ell(\boldsymbol{\xi}, D) &= \sum_{i=1}^{N} \log \int p(\boldsymbol{y}_i|\boldsymbol{b_i}) p(\boldsymbol{b_i}) d\boldsymbol{b_i} \qquad (7.1)\\
&= \sum_{i=1}^{N} [\boldsymbol{y}_i{}' \boldsymbol{x}_i' \boldsymbol{\xi} + 1_i' c(\boldsymbol{y}_i)] - \frac{N}{2} \log|D| - \frac{Nq}{2} \log(2\pi)\\
&\quad + \sum_{i=1}^{N} \log \int \exp\left[\boldsymbol{y}_i{}' \boldsymbol{z}_i' \boldsymbol{b_i} - \psi(\boldsymbol{x}_i' \boldsymbol{\xi} + \boldsymbol{z}_i' \boldsymbol{b_i}) - \frac{1}{2} \boldsymbol{b_i}{}' D^{-1} \boldsymbol{b_i}\right] d\boldsymbol{b_i}. \qquad (7.2)
\end{aligned}$$

The maximum likelihood estimates of the fixed effects $\boldsymbol{\xi}$ and covariance matrix D of the GLMM are obtained by maximizing (7.2). The problem in maximizing this likelihood, as also explained in Section 3.2.2, is the presence of N integrals over the q-dimensional random effects $\boldsymbol{b_i}$. The Gaussian variational approximation method tackles this issue by introducing an extra pair of variational parameters $(\boldsymbol{\mu}_i, \boldsymbol{\Lambda_i})$ for each subject i, $1 \leq i \leq N$, where $\boldsymbol{\mu}_i$ is a q-dimensional vector and $\boldsymbol{\Lambda_i}$ is $q \times q$ positive definite matrix. Then, new density functions $q(\boldsymbol{b_i})$ are introduced which are assumed multivariate Gaussian with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Lambda_i}$. But in principle, these densities can take any other functional form as well. In this case, the marginalized likelihood can be re-written in terms of the $q(\boldsymbol{b_i})$ densities:

$$\begin{aligned}
\ell(\boldsymbol{\xi}, D) &= \sum_{i=1}^{N} \log \int p(\boldsymbol{y}_i|\boldsymbol{b_i}) p(\boldsymbol{b_i}) d\boldsymbol{b_i}\\
&= \sum_{i=1}^{N} \log \int \frac{p(\boldsymbol{y}_i|\boldsymbol{b_i}) p(\boldsymbol{b_i})}{q(\boldsymbol{b_i})} q(\boldsymbol{b_i}) d\boldsymbol{b_i}\\
&= \sum_{i=1}^{N} \log E_{\boldsymbol{b_i} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda_i})} \left[\frac{p(\boldsymbol{y}_i|\boldsymbol{b_i}) p(\boldsymbol{b_i})}{q(\boldsymbol{b_i})}\right].
\end{aligned}$$

In this expression, $E_{\boldsymbol{b_i}\sim N(\boldsymbol{\mu}_i,\boldsymbol{\Lambda_i})}[.]$ is the expected value with respect to $\boldsymbol{b_i} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda_i})$.

By Jensen's inequality and concavity of the logarithmic function, we then have:

$$\ell(\boldsymbol{\xi}, D) \geq \sum_{i=1}^{N} E_{\boldsymbol{b_i}\sim N(\boldsymbol{\mu}_i,\boldsymbol{\Lambda_i})}\left[\log\left(\frac{p(\boldsymbol{y}_i|\boldsymbol{b_i})p(\boldsymbol{b_i})}{q(\boldsymbol{b_i})}\right)\right] = \underline{\ell}(\boldsymbol{\xi}, D, \boldsymbol{\mu}, \boldsymbol{\Lambda}),$$

where $\underline{\ell}(\boldsymbol{\xi}, D, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is the lower boundary of the loglikelihood function $\ell(\boldsymbol{\xi}, D)$, which is defined as the approximate likelihood. Alternatively, the same inequality can be derived from a Kullback-Leibler divergence point of view (Ormerod and Wand 2012; Hall, Ormerod, and Wand 2011). The accuracy of the approximate likelihood depends on the distance between $q(\boldsymbol{b_i})$ and $p(\boldsymbol{b_i}|\boldsymbol{y}_i)$, measured by the Kullback-Leibler distance. Note that, when $q(\boldsymbol{b_i}) = p(\boldsymbol{b_i}|\boldsymbol{y}_i)$, the lower bound $\underline{\ell}(\boldsymbol{\xi}, D, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is equal to $\sum_{i=1}^{N} \log \int p(\boldsymbol{y}_i|\boldsymbol{b_i})p(\boldsymbol{b_i})d\boldsymbol{b_i} = \ell(\boldsymbol{\xi}, D)$.

The idea of GVA is to approximate the computational complex posterior distribution $p(\boldsymbol{b_i}|\boldsymbol{y_i})$, because of the difficult integral, with $q(\boldsymbol{b_i})$, in such a way that the likelihoood/integrand is integrable or easier. The integral problem is not always completely removed after applying GVA. In some cases, the integral problem exists partially. However, it may still have computational advantage, although numerical approximation is required. We will see this in detail in the next section.

For the generalized linear mixed model given in (3.6), this variational lower bound of the log-likelihood simplifies to

$$\begin{aligned}\underline{\ell}(\boldsymbol{\xi}, D, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = {} & \frac{Nq}{2} - \frac{N}{2}\log|D| + \sum_{i=1}^{N}\{y_i'(\boldsymbol{x}_i'\boldsymbol{\xi} + \boldsymbol{z}_i\mu_i) + 1_i'c(y_i) - \\ & 1_i'T\left[\boldsymbol{x}_i'\boldsymbol{\xi} + \boldsymbol{z}_i\mu_i, \mathrm{diag}(\boldsymbol{z}_{\mathrm{i}}'\Lambda_{\mathrm{i}}\boldsymbol{z}_{\mathrm{i}})\right] - \frac{1}{2}\left[\log|\Lambda_i| - \mu_i'D^{-1}\mu_i - (D^{-1}\Lambda_i)\right]\},\end{aligned}$$

where

$$T\left[\mu, \sigma^2\right] = \int \psi(\mu + \sigma x)\Phi(x)dx, \tag{7.3}$$

and Φ is the standard normal cumulative distribution function.

Note that, the variational lower bound contains the original parameters $(\boldsymbol{\xi}, D)$ and additional variational parameters $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. The ML estimates are the $(\boldsymbol{\xi}, D)$ parameters obtained by maximizing the lower likelihood $\underline{\ell}(\boldsymbol{\xi}, D, \boldsymbol{\mu}, \boldsymbol{\Lambda})$.

7.2 GVA for General Frailty Models

In this section, four hierarchical models are presented for time-to-event data: the Weibull-gamma frailty model (taking in to account overdispersion in the data), the Weibull-normal random-intercept model (taking in to account a two-level hierarchy in the data), the Weibull-normal-normal random-intercepts model (taking in to account a three level hierarchy in the data), and a Weibull-gamma-normal hierarchical model, also called the Weibull-type combined model (taking both overdispersion and two level hierarchy in to account). For each of these models, a Gaussian variational approximation is derived.

7.2.1 Lower Bound for Weibull-gamma Frailty Model

Let us consider the Weibull-gamma model, as given in (6.1)–(6.2), assuming that the outcome y_{ij} follows a Weibull$(\rho, \lambda\theta_{ij}e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}})$ distribution with $\theta_{ij} \sim$ Gamma$(\alpha, 1/\alpha)$. Because of the conjugacy of the Weibull and gamma distributions, the computations to obtain the marginal likelihood simplify since the gamma frailty can be integrated out with tractable solution. The marginal likelihood can in this case be expressed as:

$$\begin{aligned}\ell(\boldsymbol{\xi}, \alpha) &= \log p(\boldsymbol{y}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}[\log(\lambda\rho) + (\rho-1)\log(y_{ij}) + \boldsymbol{x}_{ij}{}'\boldsymbol{\xi} \\ &\quad +(\alpha+1)\log(\alpha) - (\alpha+1)\log(\alpha + \lambda\rho y_{ij}^{\rho} e^{x}{}^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}})\Big]. \qquad (7.4)\end{aligned}$$

The parameters of interest are obtained by maximizing (7.4).

7.2.2 Lower Bound for Weibull-normal Random Intercept Model

Consider the Weibull-type GLMM (Weibull-normal model) where a normal random effect is used to account for clustering in the data, instead of a conjugate gamma random effect. In this case, assuming the outcome y_{ij} follows a Weibull$(\rho, \lambda\theta_{ij}e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i})$ with $b_i \sim$ Normal$(0, d)$. The marginal log-likelihood becomes:

$$\ell(\boldsymbol{\xi}, d) = \sum_{i=1}^{N}\log\int\prod_{j=1}^{n_i}\lambda\rho y_{ij}^{\rho-1}e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i}e^{-\lambda y_{ij}^{\rho}e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i}}f(b_i)db_i.$$

The N integrals of the b_i random effects do not have a tractable solution and

can lead to slow computation. Applying the Gaussian variational approximation technique leads to the Gaussian variational approximate likelihood $\underline{\ell}(\boldsymbol{\xi}, d, \boldsymbol{\mu}, \boldsymbol{\Lambda})$, as given in (7.5). This lower bound has a tractable solution, with no further requirement of an integration.

$$\begin{aligned}\underline{\ell}(\boldsymbol{\xi}, d, \mu, \Lambda) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}[\log(\lambda)+\log(\rho)+(\rho-1)\log(y_{ij})+(\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+\mu_i)- \\ &\quad \lambda y_{ij}^{\rho} e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+\mu_i+\frac{1}{2}\Lambda_i}\Big] - \frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2+\Lambda_i) + \\ &\quad \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i). \end{aligned} \tag{7.5}$$

Parameters can be obtained by maximizing this approximate likelihood. So far, we considered just one hierarchical random effect. It can be extended to two or more hierarchical random effects, e.g., the Weibull-normal-normal model. Assuming that the random effects are independent, a GVA approximation is applied to each random effect separately, leading to the lower bound:

$$\begin{aligned}\underline{\ell}(\beta, d_1, d_2, \mu, \Lambda) &= \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{n_{ij}}[\log(\lambda)+\log(\rho)+(\rho-1)\log(y_{ijk})+ \\ &\quad (\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+\mu_i+\mu_{ij}) - \lambda y_{ijk}^{\rho} e^{\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+\mu_i+\mu_{ij}+\frac{1}{2}\Lambda_i+\frac{1}{2}\Lambda_{ij}}\Big] \\ &\quad -\frac{N}{2}\log(d_1) - \sum_{i=1}^{N}\frac{1}{2d_1}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i) \\ &\quad -\frac{NM}{2}\log(d_2) - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{2d_2}(\mu_{ij}^2+\Lambda_{ij}) + \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{2}\log(\Lambda_{ij}).\end{aligned}$$

for outcome Y_{ijk} in cluster $i = 1, \ldots, N$, in subcluster $j = 1, \ldots, M_i$, measured at occasion $k = 1, \ldots, n_{ij}$. d_1 and d_2 are the variances of the first (at cluster level) and second (at subcluster level) hierarchical random effects.

7.2.3 Lower Bound for Weibull-gamma-normal Hierarchical Model

In the Weibull-type combined model (Weibull-gamma-normal model) given by equations (3.13)–(3.16) in Section 3.1.4.2, we have one additional gamma random effect,

θ_{ij}, in contrast to the Weibull-normal model. One option is to approximate the posterior density of both the gamma and the normal random effects in such a way that the integrand becomes integrable. Approximation of the posterior density of gamma random effects θ_{ij} by gamma and lognormal variational densities was attempted in which the integral problem was solved. However, it did not lead to a good approximation. An alternative way is to first integrate over the gamma random effect, as is done in partial integration. This leads to a gamma frailty model with normal random effect embedded in it and GVA is applied to the normal random effect. After integrating over the gamma random effect θ_{ij}, we have:

$$p(y_{ij}|b_i) = \frac{\lambda\rho y_{ij}^{\rho-1} e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i}\alpha^{\alpha+1}}{(\alpha+\lambda\rho y_{ij}^{\rho} e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i})^{\alpha+1}}.$$

Applying GVA leads to:

$$\begin{aligned}
\underline{\ell}(\beta, d, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \sum_{i=1}^{N} \int \log\left(\frac{p(\boldsymbol{y}_i|\boldsymbol{b_i})p(\boldsymbol{b_i})}{q(b_i)}\right) q(b_i)db_i \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n_i}(\log(\lambda)+\log(\rho)+(\rho-1)\log(y_{ij})+(\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+\mu_i)+ \\
&\quad (\alpha+1)\log(\alpha)) - \frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i) \\
&\quad - (\alpha+1)\sum_{i=1}^{N}\int \log(\alpha+\lambda\rho e^{\boldsymbol{x}_{ij}{}'\boldsymbol{\xi}+b_i})q(b_i)db_i.
\end{aligned}$$

Maximization of $\underline{\ell}(\beta, d, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ with respect to all parameters, leads to parameter estimation of (β, d). Note that, in this case a numerical approximation is still required. Approximation of the non-integrable part is done using adaptive Gaussian hermite quadrature (Liu and Pearce, 1994) which is given in Section 3.2.2.3 in general terms. The abscissae and weight are obtained from R package *statmod* (smyth, 2009).

7.2.4 Application: the Comet Assay

The proposed method was applied to the comet assay data. Three models were considered: Weibull-normal, Weibull-normal-normal and Weibull-gamma-normal. The linear predictor part of the Weibull-normal-normal model was taken the same as in Chapter 4, given in (4.7).

Estimation was done using both a GVA approximation and the numerical approximation using adaptive Gaussian quadrature in PROC NLMIXED . The latter is taken as gold standard estimate (termed 'exact estimate'). Standard software, like SAS PROC NLMIXED, does not allow for more than one hierarchical random effect level. However, for few sub-clusters, it can be modeled by using a trick. This is done by considering the subclusters as the random effect at the cluster level and specifying the same variance for the sub-clusters. For a random intercept model, k random effects for the subclusters, with k the number of subclusters, are specified at the cluster level but in an interaction with a dummy variable for the subclusters. The code is given in Appendix C.1.

The estimates using both GVA and the exact method are given in Tables 7.1 and 7.2. In terms of accuracy, the parameter estimates as well as the standard errors using GVA estimate were almost the same as the gold standard estimate for both Weibull-normal and Weibull-gamma-normal models. It was both fast and accurate. Implementation of the Weibull-normal-normal model in PROC NLMIXED had convergence problems. Therefore, we considered 10% of the data only, such that the performance of Gaussian variation approximation can still be evaluated. The result of the variational approximation is obtained in few minutes (using standard optimization method) while PROC NLMIXED took several hours for the small dataset.

7.3 GVA for Poisson Models

7.3.1 Lower Bound for Poisson-gamma, Poisson-normal, Poison-gamma-normal models

Also in the Poisson-gamma model, like in the Weibull-gamma model, the gamma random effect can be integrated out and as such the marginal likelihood does not need any approximation. If we consider a Poisson-type GLMM (Poisson-normal model), the marginal likelihood contains an intractable integral. Applying a GVA approximation results in a new lower bound likelihood with no integral problem, which is similar to that in the Weibull-normal case. It is given by:

$$\begin{aligned}\underline{\ell}(\boldsymbol{\xi}, d, \mu, \boldsymbol{\Lambda}) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\mu_i) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\mu_i+\frac{1}{2}\Lambda_i} - \log(y_{ij!})\right] \\ &\quad -\frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i).\end{aligned}$$

Table 7.1: *Comet Assay Study. Exact and GVA estimates for Weibull-normal and Weibull-gamma-normal models.*

Weibull-normal			
		Exact	GVA
Effect	Par.	Estimate(s.e.)	Estimate(s.e.)
Veh.	β_0	-13.7574(0.2270)	-13.7575(0.2270)
Low *vs.* veh.	β_1	-3.8319(0.2180)	-3.8316(0.2179)
Med.*vs.* veh.	β_2	-3.9243(0.2181)	-3.9241(0.2180)
High *vs.* veh.	β_3	-4.1268(0.2185)	-4.1266(0.2183)
Pos.C.*vs.* veh.	β_4	-2.9399(0.2653)	-2.9402(0.2652)
Weib. shape	ρ	4.3293(0.0477)	4.3293(0.0477)
Var.	d	0.1334(0.0384)	0.1333(0.0384)
-2loglik.		30476	30502.6
Duration		70 sec.	7 sec.
Weibull-gamma-normal			
		Exact	GVA
Effect	Par.	Estimate(s.e.)	Estimate(s.e.)
Veh.	β_0	-30.9295(0.7264)	-30.9292(0.7264)
Low	β_1	-42.8673(1.0005)	-42.8669(1.0004)
Med.	β_2	-43.0847(1.0045)	-43.0843(1.0040)
High	β_3	-43.5321(1.0122)	-43.5316(1.0121)
Pos.C	β_4	-40.5714(0.9768)	-40.5710(0.9767)
Weib. shape	ρ	10.7070(0.2473)	10.7070(0.2473)
Var1.	$d1$	0.9764(0.1698)	0.9764(0.1698)
OD Par.	α	0.8932(0.0463)	0.8932(0.0463)
-2loglik.		28069	28069.24
Duration		60 sec.	35 sec.

Also here, extending to more hierarchical random effects follow similarly by ap-

Table 7.2: *A subsample of Comet Assay Study. Exact and GVA estimates for Weibull-normal-normal model.*

Weibull-normal-normal			
		Exact	GVA
Effect	Par.	Estimate(s.e.)	Estimate(s.e.)
Veh.	β_0	-19.2239(0.9048)	-19.1356(0.8970)
Low *vs.* veh.	β_1	-6.9489(0.4468)	-6.9170(0.4426)
Med.*vs.* veh.	β_2	-7.1410(0.4555)	-7.1063(0.4512)
High *vs.* veh.	β_3	-7.4521(0.4670)	-7.4168(0.4627)
Pos.C.*vs.* veh.	β_4	-5.5195(0.4650)	-5.4933(0.4602)
Weib. shape	ρ	6.4192(0.2885)	6.3898(0.2861)
Var1.	$d1$	7.59e-07(0.0003)	4.81e-06(0.0002)
Var2.	$d2$	0.7184(0.1689)	0.6962(0.1644)
Duration		2 hr and 30 min.	10 min.

plying GVA to all hierarchical random effects separately.

$$\begin{aligned}\ell(\boldsymbol{\xi}, d_1, d_2, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \quad = \quad & \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{n_{ij}}\Big[y_{ijk}(\boldsymbol{x}'_{ijk}\boldsymbol{\xi} + \mu_i + \mu_{ij}) - e^{(\boldsymbol{x}'_{ijk}\boldsymbol{\xi}+\mu_i+\mu_{ij}+\frac{1}{2}\Lambda_i+\frac{1}{2}\Lambda_{ij})} \\ & - \log(y_{ijk!})\Big] - \frac{N}{2}\log(d_1) - \sum_{i=1}^{N}\frac{1}{2d_1}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i) \\ & - \frac{NM}{2}\log(d_2) - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{2d_2}(\mu_{ij}^2+\Lambda_{ij}) + \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{2}\log(\Lambda_{ij}).\end{aligned}$$

When an overdispersion gamma random effect is added to the Poisson-normal model, it leads to the Poisson-gamma-normal model (Molenberghs, Verbeke, and Demétrio 2007) given in (3.9)–(3.12), a model for repeated count data with overdispersion. The gamma random effect is first integrated out and GVA is applied to the normal random effect. In general, the Weibull and Poisson models have similar form. The

lower bound is given by:

$$\begin{aligned}\ell(\boldsymbol{\xi}, d, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}[\log((y_{ij}+\alpha-1)!) - \log((\alpha-1)!) - \log(y_{ij}!) + y_{ij}\log(\beta) \\ &\quad + y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \mu_i) - (y_{ij}+\alpha)\int \log(1+\beta e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i] \\ &\quad - \frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i).\end{aligned}$$

7.3.2 Application: the Epilepsy Study

Both the Poisson-normal and Poisson-gamma-normal are applied to the epilepsy data presented in Section 2.2. Let Y_{ij} represent the number of epileptic seizures patient i experiences during week j of the follow-up period. Also, let t_{ij} be the time-point at which Y_{ij} has been measured, $t_{ij} = 1, 2, \ldots$ until at most 27. The mean for the Poisson-gamma-normal is modelled as:

$$\log(\lambda_{ij}) = \begin{cases} (\beta_{00} + b_i) + \beta_{01}t_{ij} & \text{if placebo} \\ (\beta_{10} + b_i) + \beta_{11}t_{ij} & \text{if treated} \end{cases}$$

where the random intercept b_i is assumed to be zero-mean normally distributed with variance d. The result for the Poisson-normal and Poisson-gamma-normal models is given in Table 7.3. For Poisson-normal, it was fast and accurate and for Poisson-gamma-normal which still require numerical approximation to the resulted GVA, it was also fast and fairly accurate. Although the approximation using both methods was fast, the approximation using GVA was faster.

7.3.3 Application: the Jimma Study

Similarly, the Poisson-normal and Poisson-gamma-normal are applied to a data from Jimma Infant Survival Study, Jimma data, (Kassahun *et al* 2012). Jimma Infant Survival Study is a longitudinal study set to assess the risk factors that affect infant survival and to investigate socio-economic, maternal, and infant-rearing factors that contribute most to the child's early survival. At baseline, a total of 7969 infants enrolled in the study. The children were followed-up every two months, until the age of one year. The outcome is defined as the number of days of illness from diarrhea. Here, the effect of gender, place of residence (rural, urban and semi-urban), as well

Table 7.3: *Epilepsy Study. Exact and GVA estimates for Poisson-normal model and Poisson-gamma-normal (combined model)*

Poisson-normal			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Intercept Placebo	β_{00}	0.8179(0.1677)	0.8179(0.1675)
Slope Placebo	β_{01}	-0.0143(0.0044)	-0.0143(0.0044)
Difference in Intercept	$\beta_{10} - \beta_{00}$	-0.1703(0.2387)	-0.1704(0.2385)
Difference in Slope	$\beta_{11} - \beta_{01}$	0.0023(0.0062)	0.0023(0.0062)
Variance of RE	d	1.1568(0.1844)	1.1543(0.1839)
-2 Loglik.		-6810	-6808.87
Duration		11 sec.	4 sec.
Poisson-gamma-normal (combined)			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Intercept Placebo	β_{00}	-2.9862(0.1965)	-2.9856(0.1759)
Slope Placebo	β_{01}	-0.0248(0.0077)	-0.0248(0.0077)
Difference in Intercept	$\beta_{10} - \beta_{00}$	-0.2557(0.2500)	-0.2556(0.2498)
Difference in Slope	$\beta_{11} - \beta_{01}$	0.0130(0.0107)	0.0130(0.0107)
Var.of RE	d	1.1290(0.1850)	1.1274(0.1847)
OD par.	α	2.4640(0.2113)	2.4625(0.0324)
-2 Loglik.		-7664	-7664.17
Duration		60 sec.	50 sec.

as breast feeding behavior on the response is investigated. The linear predictor part of the model is given by:

$$\begin{aligned}\log(\lambda_{ij}) &= \beta_0 + \beta_1 R_i + \beta_2 U_i + \beta_3 G_i + \beta_4 Bf_{ij} + \beta_5 H_{ij} + \\ &\quad [\beta_6 + \beta_7 R_i + \beta_8 U_i + \beta_9 G_i + \beta_1 0 Bf_{ij} + \beta_1 1 H_i].t_{ij} + b_i,\end{aligned}$$

where R_i and U_i are dummy variables for place of residence corresponding to rural and urban areas, with using semi-urban areas as a reference. G_i is the Gender indicator, t_{ij} is the time point at which the j^{th} measurement is taken for the i^{th} infant. Bf_{ij} denotes whether the i^{th} infant is breast fed or not at time j (months). The random intercept $b_i \sim N(0, d)$. The result for Poisson-normal and Poisson-gamma-normal is given in Table 7.4. The gain in computational time is clearly seen in this analysis.

Table 7.4: *Jimma Study. Exact and GVA estimates for Poisson-normal model and Poisson-gamma-normal model (combined model).*

Poisson-normal			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Intercept	β_0	1.6529(0.1039)	1.6938(0.1028)
Rural	β_1	0.4627(0.0636)	0.4561(0.0615)
Urban	β_2	-0.2654(0.0832)	-0.2657(0.0806)
Time	β_3	-0.1268(0.0089)	-0.1269(0.0089)
Gender	β_4	0.1017(0.0440)	0.1092(0.0431)
Breast feeding	β_5	-1.5016(0.0845)	-1.5034(0.0844)
Help	β_6	-2.9504(0.0242)	-2.9533(0.0242)
Slope Rural	β_7	-0.0102(0.0030)	-0.0102(0.0030)
Slope Urban	β_8	0.0531(0.0040)	0.0531(0.0040)
Slope Gender	β_9	-0.0059(0.0023)	-0.0059(0.0023)
Slope Breast feeding	β_{10}	0.1430(0.0084)	0.1431(0.0084)
Slope Help	β_{11}	0.1666(0.0028)	0.1666(0.0028)
log. Variance of RE	$\log(d)$	1.3473(0.0273)	1.2616(0.0177)
Duration		45 min.	3 min.
Poisson-gamma-normal (Combined)			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Intercept	β_0	5.7516(0.4570)	5.7120(0.4495)
Rural	β_1	-0.0697(0.1230)	-0.0681(0.1195)
Urban	β_2	-0.2946(0.1599)	-0.2908(0.1551)
Time	β_3	-0.3282(0.0506)	-0.3255(0.0498)
Gender	β_4	0.2426(0.1039)	0.2389(0.1010)
Breast feeding	β_5	-3.1524(0.4151)	-3.1199(0.4079)
Help	β_6	-6.1489(0.1896)	-6.0022(0.1882)
Slope Rural	β_7	0.0179(0.0158)	0.0178(0.0156)
Slope Urban	β_8	0.0792(0.0201)	0.0778(0.0198)
Slope Gender	β_9	0.0064(0.0129)	0.0063(0.0127)
Slope Breast feeding	β_{10}	0.3209(0.0453)	0.3181(0.0446)
Slope Help	β_{11}	0.3448(0.0219)	0.3396(0.0217)
log Var.of RE	$\log(d)$	1.0432(0.0514)	0.8925(0.0525)
log OD par.	$\log(\rho)$	-2.2590(0.0199)	-2.2802(0.0175)
Duration		4 hr and 40 min	5 min.

7.4 GVA for Logistic Models

7.4.1 Lower Bound for Logistic-normal and Logistic-beta-normal Models

For the logistic-normal model, we have:

$$\begin{aligned} Y_{ij} &\sim \text{bernoulli}(\pi_{ij}), && (7.6)\\ \text{logit}(\pi_{ij}) &= \boldsymbol{x}'_{ij}\boldsymbol{\xi} + b_i, && (7.7)\\ b_i &\sim \text{Normal}(0, d). && (7.8) \end{aligned}$$

Applying GVA results in the lower bound $\underline{l}(\boldsymbol{\xi}, d, \mu, \Lambda)$. Unlike in the Poisson-normal and Weibull-normal models, the GVA likelihood still has a non-tractable likelihood:

$$\begin{aligned} \underline{\ell}(\boldsymbol{\xi}, d, \mu, \Lambda) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \mu_i) - \int \log(1 + e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i\right] \\ &\quad -\frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2 + \Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i). \end{aligned}$$

For the Bernoulli-type combined model (logistic-beta-normal models) given in (3.22)–(3.23), the lower bound is given by:

$$\begin{aligned} \underline{\ell}(\boldsymbol{\xi}, d, \alpha, \beta, \mu, \Lambda) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}\Bigg[y_{ij}\log(\alpha) - \log(\alpha+\beta) - \int \log(1 + e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i + \\ &\quad y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \mu_i) + (1-y_{ij})\int \log(\alpha+\beta+\beta e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i\Bigg] \\ &\quad -\frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2 + \Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i). \end{aligned}$$

For identifiability reasons, we fix $\alpha/\beta = c$. The lower bound is then given by:

$$\begin{aligned}\underline{\ell}(\boldsymbol{\xi}, d, c, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}\Big[-\log(1+c) - \int \log(1+e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i + \\ &\quad y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\mu_i) + (1-y_{ij})\int \log(1+c+ce^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+b_i})q(b_i)db_i\Big] \\ &\quad - \frac{N}{2}\log(d) - \sum_{i=1}^{N}\frac{1}{2d}(\mu_i^2+\Lambda_i) + \sum_{i=1}^{N}\frac{1}{2}\log(\Lambda_i).\end{aligned}$$

We see that the GVA for the Weibull-gamma-normal, Poisson-gamma-normal and logistic models, have similar algebraic forms, which still needs numerical approximation.

7.4.2 Application: the EG Study

We applied the proposed method to the EG data. The model is given by:

$$\text{logit}\left[P(y_{ij}=1|\boldsymbol{\xi}, b_i)\right] = \beta_0 C_{ij} + \beta_1 L_{ij} + \beta_2 M_{ij} + \beta_3 H_{ij} + b_i,$$

where C_{ij} is an indicator for the control group and L_{ij}, M_{ij}, and H_{ij} are the indicators for low, medium, and high dose groups, respectively. The random intercept b_i corresponds to the animal-specific random effect which is assumed to be zero-mean normally distributed with variance d. The result is presented in Table 7.5. The performance of the GVA in terms of the accuracy of the parameter estimates as well as the standard error for the logistic models was slightly lower as compared to the Weibull and Poisson models.

Table 7.5: *Ethylene Glycol Study. Exact and GVA estimates for logistic-normal and logistic-beta-normal (combined model)*

logistic-normal			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Control	β_0	-6.2344(0.8860)	-6.1592(0.8543)
Low	β_1	-3.8615(0.5420)	-3.8011(0.5190)
Medium	$\beta_2 - \beta_{00}$	-1.7370(0.4332)	-1.6984(0.4121)
High	$\beta_3 - \beta_{01}$	1.5695(0.4693)	1.5274(0.4466)
Var. of RE	d	3.9988(1.0977)	3.5808(0.9360)
Duration		18 sec.	6 sec.
logistic-beta-normal			
		Exact	GVA
Effect	Parameter	Estimate(s.e.)	Estimate(s.e.)
Control	β_0	-6.2344(0.8860)	-6.1592(0.8543)
Low	β_1	-3.8615(0.5420)	-3.8011(0.5190)
Medium	$\beta_2 - \beta_{00}$	-1.7370(0.4332)	-1.6984(0.4121)
High	$\beta_3 - \beta_{01}$	1.5695(0.4693)	1.5274(0.4466)
Var. of RE	d	1.3860(0.2745)	1.2756(0.2614)
OD par.	β/α	1.30e-07(0.00011)	2.68e-09(0.00001)
Duration		35 sec.	40 sec.

Table 7.6: *Overview of computational efficiency, duration for convergence (h:m:s).*

Models	Datasets	Exact	GVA
Weibull-normal	Comet	0:01:10	0:00:07
Weibull-gamma-normal	Comet	0:01:00	0:00:35
Poisson-normal	Epilepsy	0:00:11	0:00:04
Poisson-gamma-normal	Epilepsy	0:01:00	0:00:50
Poisson-normal	Jimma	0:45:00	0:03:00
Poisson-gamma-normal	Jimma	4:40:00	0:05:00
logistic-normal	EG	0:00:18	0:00:06
logistic-beta-normal	EG	0:00:35	0:00:40

7.5 Discussion and Conclusion

Generalized linear mixed models often involve intractable integrals. Different approximating techniques exist, which can be broadly categorized as approximating the integrand, the data or the integral itself. Gaussian variational approximation approximate the integrand by introducing a set of variational densities (to the posterior densities) in such a way that their evaluation is tractable. It is applicable for both Bayesian and likelihood paradigms. In this chapter, we focused on the likelihood framework by approximating the posterior density of the normal random effect (by a set of normal densities). We considered three families of GLMM models: 1) the Weibull models: Weibull-normal, Weibull-normal-normal, and Weibull-gamma-normal; 2) the Poisson models: Poisson-normal, Poisson-normal-normal and Poisson-gamma-normal; 3) the logistic models: logistic-normal and logistic-beta-normal.

The GVA approximation was applied to the comet assay data for Weibull models, epilepsy data for Poisson models and EG data for logistic model. Estimation using adaptive numerical Gaussian quadrature in the SAS Procedure NLMIXED was taken as gold standard. For Weibull-normal and Poisson-normal, estimation using GVA was faster and very accurate (in contrast with the exact estimate). For models with higher hierarchical random effect (Weibull-normal-normal), this is not accomodated in conventional software, such as the SAS procedure NLMIXED. It is only possible with the use of some modeling tricks for cases of small number of sub-clusters. Turning to the comet data, we were having problems in convergence. Thus, we were forced to deal with the reduced data, yet it was taking very long time to converge. Estimation using GVA was much faster and fairly accurate. Considering overdispersed hierarchical models (Weibull-gamma-normal, Poisson-gamma-normal, and logistic models), applying GVA approximation still requires numerical approximation. It was also fast and fairly accurate for the parameters of interest.

In general, GVA can be a good approximation technique especially when the numerical approximation using standard software fails, is very restrictive, takes a long time, exhibits problems in convergency, or when it does not allow to accommodate such features, for instance, when we have more than two hierarchical levels where and when we have higher dimension of random effect. Table 7.6 gives an overview of the computational efficiency in terms of time to convergence.

Chapter 8

Discussions and Concluding Remarks

This first part of the thesis is partly motivated by comet assays, a toxicology study commonly encountered in preclinical research to assess DNA damage. Nowadays, the assay has gained widespread use in various areas and has emerged as a standard tool in the pharmaceutical industry for the assessment of the safety of potential new drugs. It has a higher-order hierarchical representation. In essence, the comet assay represents a hierarchical design with animals nested within doses, a number of slides per animals and several cells measured per slide. In general, comet measures from an animal are clearly not normally distributed but are rather asymmetric, skewed, bi- or multimodal, a mixture of different distributions, etc. The different type of measures used for the quantification of the DNA damage also exihibits a multivariate nature and three measures are commonly used: the tail migration (i.e., tail length), percentage of tail intensity, and tail moment. While such data consist of non-Gaussian outcomes in a multi-level hierarchical structure, traditional analyses typically completely or partly ignore some of the ingredients that are useful in modeling and that could have a major impact on the inference drawn.

In this first part, we have explored some statistical models for hierarchically structured and overdispersed outcomes for comet assay data and propose an alternative estimation technique for the combined model in general.

In the univariate analysis, a flexible modeling approach for hierarchically clustered and overdispersed non-Gaussian outcomes for comet assays is proposed that exihibits

a full hierarchical structure, an appropriate distribution, and possible overdispersion using a combined model (Molenberghs *et al* 2010). In this approach, while a conjugate gamma random effect is used for the overdispersion random effect, both gamma and normal random effects are considered for the hierarchical random effect, the more conventional models with either the overdispersion, or just one hierarchical random effect being submodels. It provided a wide choice of models to select from. The use of more elaborate model with overdispersion and hierarchical structure improved the fit for one response. In this approach, the observation-specific gamma random effects within a cluster are assumed to be independent. It could also further be explored to see whether the result improves by considering correlated conjugate random effects with different covariance structure.

In the combined model, a conjugate distribution is assumed for the overdispersion random effect. In this thesis, alternative distributions are considered. A discrete distributions that eventually lead to finite mixture models as an alternative way of accounting for overdispersion. Zero-inflated models are also explored to account for the excess zeros. A mixture of conjugate distributions for the overdispersion random effects is also considered, where deviations from a single conjugate distribution are allowed for while the property of conjugacy can still be employed to ease computations. In addition, it would have been possible to consider other distributions, like a normal random effect added in the linear predictor, which will then have a log-normal distribution in the multiplicative factor. Some assessment of misspecification of the distribution of the overdispersion random effect and thereof the impact on the inference drawn could further be investigated.

In multivariate analysis, the hierarchically clustered and overdispersed non-Gaussian outcomes are jointly analyzed using a multivariate normal distribution. It allowed to draw overall inferences using all outcomes and to capture the association among them. Using correlated hierarchical random effects is less restrictive compared to shared random effects. But it still assumes that the association is entirely induced by the hierarchical random effect. If the two random effect are uncorrelated then there is no association between the two outcomes. To accomodate deviations from this assumption, it would have been sensible to explore whether the association is through the observation specific overdispersion random effect or both. In that case, one may opt for multivariate extensions of the gamma model. In the case of a bivariate setting as in the comet data, it would also be possible to use copulas.

Most of the statistical models considered are using the combined model of Molenberghs *et al* (2010) or extensions thereof. In general, the main difficulty with this

kind of models is the computationally complex estimation due to the intractable multivariate integrals, as is the case for generalized linear mixed models that involve such integrals with no analytic expression. In this thesis, the use of Gaussian variational approximation is explored as an alternative estimation technique. It approximates the integrand by introducing a set of variational densities in such a way that their evaluation is tractable. It is quite useful in the cases where there is computational difficulty using standard estimation or when computational time would be excessive otherwise.

In this thesis, we have discussed some modeling issues that could distort the inferences drawn if they are not properly addressed in modeling. We tried to accommodate them by proposing a flexible modeling framework. The development of such elaborate models is quite useful because even when the more elaborate model does not provide a substantially improved fit, nor alters the inferences drawn, the development is still very useful because it provides further confidence, by way of model specification assessment.

Part II

Assessment of Type I Error Rate Associated with Dose Group Switching in Clinical Trials with Incomplete Data

Chapter 9

Introduction and Fundamental Concepts of Incomplete data and Multiplicity

In clinical trials, although the primary focus of many trials is on a specific time of measurement - usually the last, the outcome of interest is recorded in a longitudinal fashion, and missingness in general, and dropout in particular is a common occurrence (Molenberghs and Kenward, 2007). Missing data has a potential to lead to biased results and there can be a severe loss of power if the proportion of incompleteness is high thus, endangering the credibility of the inference drawn. The effect could be more in dropout, where subjects are observed without interruption from the beginning of a study to a given point in time when they drop out and do not return to the study. Although model formulation and manipulation could be simplified, the causes behind it could be more problematic. This may stem from lack of efficacy, or from potentially serious treatment-related side effects that can have impact on the inference (i.e. significance of the treatment effect). On the other hand, an intermittently missing endpoint value may be due more plausibly to the patient skipping for practical or administrative reasons, to measurement equipment failure, etc. As such, the impact could be less in terms of the bias. However, the effect of missingness cannot be

underestimated, it needs to be handled with care in order to draw valid inference.

While analysing incomplete data, in order to obtain valid inference, one needs to ask what the response of the subject would have shown had they remained in the study (Molenberghs and Kenward, 2007). Considering nature of the 'missing data mechanism' on that matter is important. This is not under the control of the investigators, and only assumptions about it are made. Based on the assumption, an appropriate method for handling missing data is applied. There are various methods, ranging from simple methods to multiple imputation and likelihood ignorable methods. However, the validity of the analyses will always depend on whether these assumptions hold for the particular case. It is always better to minimize the missingness to the level possible so as to minimize the impact. The impact increases with the increase in the percentage of missingness in the data. For example, the effect on the power of a test is higher when the dropout is higher.

Some clinical trials allow for a data-driven adaptation when the missingness/-dropout rate is very high. For example, if the dropout rate in one of the treatment arms is very high, that treatment group is dropped and another treatment arm with less missing data used instead. That is, one allows for switching treatment comparison. However, reliability of the inference drawn is questioned as it can have an impact on the type I error rate. It is believed by some that it inflates the type I error rate and may need multiplicity adjustment. The focus here is to assess the impact of such data-driven adaptation. In particular, the impact of switching treatment comparison at the end of a trial on the type I error rate. The problem at hand differs from the common multiple testing problem. In a way, in multiplicity, analyses are done with the presumption that attention will focus on the strongest differences among all comparisons that are made but, in this case, no multiple comparisons are done. Rather, one comparison is selected at the end of the study based on dropout rate. But still the question remains whether such data driven adaptations will have impact on the type I error rate.

In this chapter, we present the methods for incomplete data and the basic concepts of type I error rate and multiplicity. Common terminology for missing data will be introduced and fundamental concepts of missing data, different modeling frameworks, and the missing mechanisms will be reviewed and methodologies for handling them (simple imputation methods, multiple imputation, and ignorable likelihood methods) will be presented.

In Chapter 10, the impact of data driven adaptation on type I error rate will be assessed. A simulation experiment will be set up to assess the type I error rate

inflation associated with switching dose group as a function of drop out rate at the end of the study, where the primary analysis is in terms of a longitudinal outcome. The work is inspired by a clinical trial in Alzheimer's disease. The type I error rate will be assessed under a number of scenarios, in terms of different correlations between efficacy and tolerance, different missingness mechanisms, and different probabilities of switching. A collection of parameter values will also be used to assess sensitivity of the analysis.

9.1 Incomplete Data

The following terminology is based on the standard framework of Rubin (1976) and Little and Rubin (2002). Let the random variable Y_{ij} denote the response for subject $i = 1, \ldots, N$ at occasion $j = 1, \ldots, n_i$. The outcomes are grouped into a vector $\boldsymbol{Y_i} = (Y_{i1}, \ldots, Y_{in_i})$. Define further a vector of missingness indicators $\boldsymbol{R_i} = (R_{i1}, \ldots, R_{in_i})$ with $R_{ij} = 1$ if Y_{ij} is observed and 0 otherwise. The set of measurements, along with the missingness indicators, ($\boldsymbol{Y_i}$,$\boldsymbol{R_i}$), comprise what is called the full data. Typically, the vector $\boldsymbol{Y_i}$ is divided into observed ($\boldsymbol{Y_i^o}$) and missing ($\boldsymbol{Y_i^m}$) components, respectively. For incomplete data, only ($\boldsymbol{Y_i^o}, \boldsymbol{R_i}$) is available.

Two types of missingness exist based on the pattern of missingness: *monotone* and *non-monotone*. When the missingness is monotone or of a dropout nature, the unobserved measurement within the longitudinal series all occur after a particular measurement occasion, and in that sense, the subject is said to have "dropped out" of the study. In such cases, the missingness indicator $\boldsymbol{R_i}$ consists of a very particular form, with all R_{ij} equal to one up to a particular time point j and zero thereafter. This structure allows the missingness indicators in R_i to be collapsed into a single variate, D_i, defined as $D_i = 1 + \sum_{j=1}^{n_i} R_{ij}$ denoting the time point at which subject i drops out. Non-monotone missingness on the other hand, occurs when missing values arise intermittently within the series.

9.2 Modeling Frameworks

In modeling missing data, one would need to consider the full data density $f(\boldsymbol{y}_i, \boldsymbol{r_i} | \boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}, \boldsymbol{\psi})$, where $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are the design matrices for fixed and random effects, respectively, and the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes. The parameter $\boldsymbol{\theta}$ is composed of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, in

order, are the fixed effect and covariance parameters. This full density function can be factored in different ways, each leading to a different framework. Under a selection model framework (Rubin 1976, Little and Rubin 2002), the joint distribution is factored into a marginal density of the measurement process and a conditional model for missingness process given the outcomes, that is,

$$f(\boldsymbol{y_i}, \boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y_i}|\boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{\psi}). \tag{9.1}$$

Selection models are a primary choice if one is interested in the marginal effect, $\boldsymbol{\theta}$, of the independent variables on the response. Alternatively, one can consider so-called pattern-mixture models (Little 1993, 1994a, Molenberghs *et al* 1997), using the reversed factorization:

$$f(\boldsymbol{y_i}, \boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y_i}|\boldsymbol{r_i}, \boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{\psi}).$$

This density can be seen as a mixture of different populations, each of which is defined conditionally on the observed pattern of missingness. The parameters $\boldsymbol{\theta}$ then denote pattern-specific effects of the independent variables on the response. Instead of using the selection or pattern-mixture model frameworks, the measurement and the dropout process can be jointly modeled using a shared-parameter model (Wu and Carroll 1988, Wu and Bailey 1989). In such a model the measurement and dropout process are assumed to be independent, conditional upon a certain set of shared parameters. It is given by:

$$f(\boldsymbol{y_i}, \boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{b_i}, \boldsymbol{z_i}, \boldsymbol{\psi}).$$

Here, $\boldsymbol{b_i}$ are the shared parameters, often considered to be random effects and following a specific parametric distribution. Further, $\boldsymbol{\theta}$ denotes the effects of the covariates, conditional on the random effects.

9.3 Missing Data Mechanisms

In order to obtain valid inferences from incomplete data, one needs to consider the nature of the "missing data mechanism". Normally, the missing data mechanism is not under the control of the investigators; consequently, it is often not well understood. Instead, assumptions are made about the missing data mechanism, and the validity

of the analyses will depend on whether these assumptions hold for the data at hand.

Rubin (1976) developed a taxonomy to classify the missingness process based on its dependence on the measurement process. This classification is based on the second term of (9.1). Upon partitioning of the response vector into its observed and missing components, it can be expressed as $f(\boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{\psi}) = f(\boldsymbol{r_i}|\boldsymbol{x_i}, \boldsymbol{y_i^o}, \boldsymbol{y_i^m}, \boldsymbol{\psi})$. When there is independence on the measurement and missingness process, conditionally of the covariates, the mechanism is missing completely at random (MCAR). A less strict assumption would be missing at random (MAR), for which the missingness may depend on the observed outcomes and covariates but not on the unobserved outcomes. If the cause of missing data is neither MCAR nor MAR, the data is missing not at random (MNAR).

Consider a longitudinal clinical trial to assess the efficacy of a new treatment for a particular disease or condition. If a patient drops out for a reason not related to treatment effect, like due to weather condition, family issue, this most probably falls within the category of MCAR, since the missingness process and the outcome are independent. On the other hand, if a patient missed a visit because in previous visits his condition stabilized and he is convinced that continuing the visits to the hospital are of no value or if the result of the previous visits was discouraging that he decided to stay at home, then the nature of the missingness is related to the previously observed outcome, and the most plausible process is MAR. In general, if dropping out is known to be unrelated to current health conditions, an MAR assumption for the missing values seems justified; however, if dropping out is related to current health conditions then the MAR assumption is not justified, and the missing data are likely MNAR.

9.4 A Model for Continuous Longitudinal Data

Here, we consider a model for longitudinal data in a Gaussian setting. The most widely used methodology for continuous longitudinal data within the likelihood framework is the general linear mixed-effects model (Verbeke and Molenberghs 2000), which takes the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{9.2}$$

where $\boldsymbol{Y}_i$ is the n_i-dimensional response vector for subject $i = 1, \ldots, N$. $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are, respectively, $(n_i \times p)$ and $(n_i \times q)$ known design matrices. $\boldsymbol{\beta}$ is the p-dimensional

vector containing the fixed effects. $\boldsymbol{b}_i \sim N(\mathbf{0}, D)$ is the q-dimensional vector containing the random effects. $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ are the within-subject random errors. D and $\boldsymbol{\Sigma}_i$ are general covariance matrices of size $(q \times q)$ and $(n_i \times n_i)$, respectively.

From (9.2), conditional on the random effect $\boldsymbol{b}_i$, $\boldsymbol{Y}_i$ is normally distributed with mean vector $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ and with covariance matrix $\boldsymbol{\Sigma}_i$. Upon integration over the random effects, the resulting marginal model for the response can be expressed as:

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i D \boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i).$$

As shown by Verbeke and Molenberghs (2000), the random-effects model implies a simple marginal model in the linear mixed model case. The expectation $\boldsymbol{X}_i\boldsymbol{\beta}$ follows either by (1) marginalizing over the random effects or by (2) conditioning on the random-effects vector $\boldsymbol{b}_i = 0$. Thus, the fixed-effects parameters $\boldsymbol{\beta}$ have both a marginal and a hierarchical model interpretation.

9.5 Methodology for Incomplete Data

In this section, some of the methods commonly used for handling incomplete data in longitudinal data analysis in a Gaussian setting are presented: simple ad-hoc methods, imputation methods, and maximum likelihood estimation methods. We also discuss the assumptions that need to be made about the missing data mechanism for each method to yield valid inferences.

9.5.1 Simple Methods

Two simple, common methods to analyze incomplete data are complete case analysis, which discards subjects with incomplete sequences, and simple imputation. Last observation carried forward, for which the last observed measurement is substituted for values at later points in time that are not observed, is among the commonly used simple imputation methods. Until recently, clinical trial practice has put a strong emphasis on such simple methods. Some of the claimed advantages include computational simplicity, no need for a full longitudinal model (e.g., when the research question is in terms of the last observed measurement occasion only) and, for LOCF, compatibility with the Intention-to-Treat (ITT) principle, since data on all patients randomized can be used.

9.5.1.1 Complete Case Analysis

A complete case analysis (CC) includes only those cases for analysis, for which all measurements (covariates and outcomes) were recorded (Verbeke and Molenberghs, 2000; Little and Rubin, 2002; Molenberghs and Verbeke, 2005). This method has the advantage of simplicity. But it is an inefficient use of information, with adverse effects on precision and power. Further, such an analysis will only be representative for patients who remain on study and have complete data. In addition, severe bias can result when the missingness mechanism is MAR. For example, if the completers are the ones with a better result, then they are not representative of the population at large and would overestimate the effect. This method is valid only under MCAR.

9.5.1.2 Last observation Carried forward

Last observation carried forward (LOCF) is a common single imputation method where the most recent observation replaces any subsequent missing ones. It can be applied to both monotone and non-monotone missingness. The idea of LOCF is based on the very strong and unrealistic assumption that a subject's measure stays at the same level until the end of the trial or during the period they are unobserved in the case of intermittent missingness. In most clinical trial settings, the assumption that a patient's condition would remain at the response level is questionable as study effects, placebo effects, and natural time evolution also influence outcomes. Molenberghs and Kenward (2007) showed, using hypothetical data, that, even under the unrealistically strong assumption of MCAR, while CC produces unbiased estimates, the bias in the LOCF estimator does not vanish, and can even induce an apparent treatment effect when there is none. Under MAR, they showed that both can be biased and bias can go in either direction.

9.5.2 Multiple Imputation

A widely used approach for handling incomplete data is using some form of imputation. The basic idea behind imputation is simple: substitute or fill in the values that were not recorded with the imputed values. Methods that impute or fill in the missing values have the advantage that, unlike CC, the information from the observed values in the incomplete cases is retained and once a filled-in data set has been constructed, standard methodology for complete data can be applied. However, single imputation methods, creating only a single filled-in data set, fail to acknowledge the uncertainty

inherent in the imputation of the unobserved responses. Multiple imputation (MI) circumvents this difficulty. MI was formally introduced by Rubin (1978). The key idea of the procedure is to first replace each missing value with a set of M plausible values drawn from the conditional distribution of the unobserved values, given the observed ones. This conditional distribution represents the uncertainty about the right value to impute. In this way, M imputed data sets are generated, which are then analyzed using standard complete data methods. Finally, the results from the M analyses have to be combined into a single inference by means of the method laid out in Rubin (1978). In its basic form, multiple imputation requires the missingness mechanism to be MAR.

9.5.3 Maximum Likelihood Estimation

We have presented maximum likelihood in the first part of the thesis (Chapter 3) in general terms and here, we present the concept in the context of missing data. When data are incomplete and under a selection model framework, subject i's observed-data likelihood contribution takes the form:

$$L_i = \int f(\boldsymbol{y}_i|\boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{y_i^o}, \boldsymbol{y_i^m}, \boldsymbol{\psi}) d\boldsymbol{y_i^m}. \tag{9.3}$$

In general, (9.3) does not simplify, but under MCAR (or MAR), we obtain respectively:

$$L_i = f(\boldsymbol{y_i^o}|\boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{\psi}), \tag{9.4}$$

or

$$L_i = f(\boldsymbol{y_i^o}|\boldsymbol{\theta}) f(\boldsymbol{r_i}|\boldsymbol{y_i^o}, \boldsymbol{\psi}). \tag{9.5}$$

Hence, likelihood inferences for the measurement model parameters $\boldsymbol{\theta}$ can be made without explicitly formulating the missing data mechanism, provided the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are disjoint, that is, their joint parameter space is the cartesian product of the two component parameter spaces (Rubin 1976). It is precisely this result which makes so-called direct likelihood analyses valid under MCAR and MAR.

9.6 Type I Error Rate and Multiplicity

9.6.1 Type I Error Rate

In statistical inference, type I error and type II error are fundamental concepts. Any statistical hypothesis test has a probability of making type I and type II errors. A type I error refers to an incorrect rejection of a true null hypothesis whereas a type II error is a failure to reject a false null hypothesis. A type I error leads one to conclude that a relationship exists when in reality there is none, for example, that a certain treatment cures a disease when in reality it does not. A small nominal level of significance, usually 5%, is allowed for a result to be significant when there are no relationships in the population. However, the actual error rate in some designs may be more than the allowed error rate and this phenomenon is called inflation of the overall Type I error rate. For example, the overall Type I error rate is higher when investigates a lot of effects in the data. In general, type I error rate is not always controlled. For some designs, it is questioned and one of these will be discussed in the next chapter.

9.6.2 Multiplicity

Multiple comparisons arise when a statistical analysis encompasses a number of formal comparisons, with the presumption that attention will focus on the strongest differences among all comparisons made. The term 'comparison' refers to comparison of two groups, such as a treatment group and a control group. Failure to compensate for multiple comparisons can have important real-world consequences. Suppose we consider the efficacy of a drug in terms of the reduction of any one of a number of disease symptoms. As more symptoms are considered, it becomes more likely that the drug will appear to be an improvement over existing drugs in terms of at least one symptom. As the number of comparisons increases, it becomes more likely that the groups being compared will appear to differ in terms of at least one attribute. The confidence that a result will generalize to independent data should generally be weaker if it is observed as part of an analysis that involves multiple comparisons, rather than an analysis that involves only a single comparison. In this respect, multiplicity adjustment have been considered important. Several statistical techniques have been developed, allowing significance levels for single and multiple comparisons to be directly compared. These techniques generally require a stronger level of evidence to be observed in order for an individual comparison to be deemed "significant", so as to

compensate for the number of inferences being made. Some of these are Bonferroni's correction, and False discovery rate using Benjamini and Hochberg method to correct for multiple comparisons (Benjamini and Hochberg 995).

Chapter 10

Assessment of Type I Error Rate Associated with Dose Group Switching in a Longitudinal Alzheimer Trial

In clinical trials, it is not uncommon to modify trial and/or statistical procedures during conduct, based on review of interim data, or even at the end of the study. Such adaptive designs have been in use for quite a while now. Procedural changes may also be implemented at the end of the study; this is of interest here. Adaptation oftentimes reflects medical practice, and may be regarded as ethical conduct with respect to both efficacy and tolerance of the experimental treatment. However, it is a concern whether the p-value or the confidence interval associated with the treatment effect after modification can be reliably and correctly compared to the nominal α level (Chang 2008, Chow and Chang 2008, Wang, Wu, and Tsai 2008).

The objective here is to examine key operating characteristics of a clinical trial design with data-driven adaptation, when the primary analysis is based on a longitudinal outcome. The assessment of the type I error rate inflation associated with adaptation of a trial by switching dose groups at the end of the study is scrutinized in particular. This work is motivated by a clinical trial in Alzheimer's disease. It is believed by some that switching dose groups may lead to an inflated Type I error rate

and thus the significance level needs to be adjusted. On the other hand, it may not be applicable to some trials, including the Alzheimer's disease trial considered here. The Type I error rate can be lower for such trials, thus not inflating the overall Type I error rate, when the primary analysis is modified by switching doses based on high dropout rate in the high dose arm.

In this chapter, the Type I error rate inflation related with switching the dose level in the primary analysis is assessed by comparing the estimated Type I error rate with switching doses and without switching, i.e., adhering to the pre-specified comparison. These results were summarized in Ghebretinsae *et al* (2013).

Section 10.1 offers some theoretical background on the inflation of Type I error rate associated with switching treatment comparison. Section 10.2 describes the design of the Alzheimer clinical trial considered and the statistical model employed. Section 10.3 presents the design of the simulation study, the results of which are described in Section 10.4. Finally, the findings are discussed in Section 10.5.

10.1 Theoretical Background on Changes in the Primary Analysis and Their Effects on Type I Error Rate

We consider a clinical trial with a primary endpoint of which the primary analysis can be performed using two different comparisons. A priori, the intention is to use comparison 1 (High dose *versus* Placebo), but one can switch to comparison 2 (Low dose *versus* Placebo). The comparison is selected at the end of the study and the decision for switching is driven by the data collected in the study. That is, depending on the drop-out rate in the high-dose arm one of the possible comparisons is chosen.

Whether switching a comparison is allowed or not (in the latter case, one sticks to the first comparison), the primary analysis is performed at a pre-specified α level. Evidently, a key question is whether this switching strategy inflates the Type I error rate. To address this question, denote H as the test statistic used in the decision for switching rule. If $H < h$, the primary analysis will be performed using the first comparison, otherwise the second comparison is employed. In addition, let t_1 and c_1 represent the test statistic and critical value for comparison 1, respectively. In this case, if $t_1 \geq c_1$, the primary analysis becomes significant, based on comparison 1. Note that, under the hypothesis of no treatment difference, the error rate is approximately

the nominal rate $[P(t_1 \geq c_1) \leq \alpha]$. Similarly, take t_2 and c_2 to be the test statistic and critical value for comparison 2. Then, as before, the error rate is approximately the nominal rate $[P(t_2 \geq c_2) \leq \alpha]$, under the hypothesis of no treatment difference.

If the trial does not allow for switching, then the comparison is entirely based on the high dose and placebo arms. Evidently, then there is neither a multiple testing problem nor error rate inflation. If switching is allowed, then the type I error rate is given by:

$$[P(H < h)P(t_1 \geq c_1|H < h)] + [P(H \geq h)P(t_2 \geq c_2|H \geq h)].$$

Because low dose is not part of the switching criterion, the test of comparison 2 does not depend on the switching process, i.e., t_2 and H are independent. Controlling the type I error rate is equivalent to having $P(t_1 \geq c_1|H < h) \leq \alpha$. If the decision for switching and testing high dose *versus* placebo comparison are independent then the type I is protected, i.e., $P(t_1 \geq c_1|H < h) = P(t_1 \geq c_1) \leq \alpha$. However, in general, the Type I error rate may not be preserved at the α level. The amount of Type I error rate inflation is likely to increase with the increase in correlation between the decision for switching and first comparison. This will be scrutinized in the current chapter.

10.2 A Clinical Trial in Alzheimer's Disease

Alzheimer's disease (AD) is a neurodegenerative disorder causing progressive decline in memory and other aspects of cognition. The average duration from onset of symptoms to nursing home placement is 5 to 7 years and from symptom onset to death is 7 to 9 years (Figure 10.1). Although 6-month trials are still standard in regulatory guidelines for AD trials, 18 month long randomized placebo controlled trials are very common. Discontinuation rates for any cause, including death, vary across long-term AD trials and range between 20% and 40%. Thus, it is critical to account for the high discontinuation rates in the design and analysis of AD studies.

A phase III clinical trial was designed for patients who were at least 55 years old to assess the effect of an experimental treatment (ET) compared to placebo on AD progression using co-primary endpoints that include both cognitive and functional measures. This study was a multi-site (176 sites), randomized, double-blind, placebo-controlled, Phase III study of 1500 patients to compare 3 treatments: (i) high dose of the experimental treatment, (ii) low dose of the experimental treatment, and (iii)

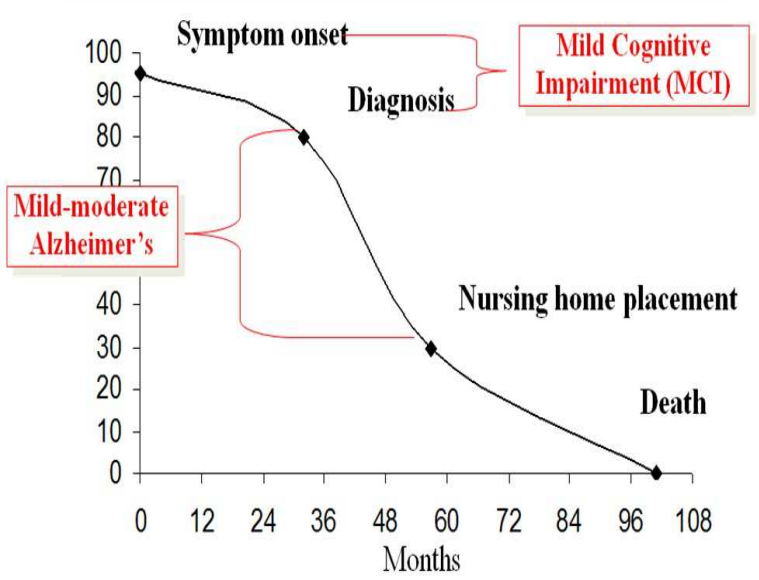


Figure 10.1: *Alzheimer's disease. Onset and expected decline in memory and cognition.*

placebo. The co-primary endpoints for cognition and function were assessed on all patients at baseline (prior to start of treatment) and six post-baseline visits (at weeks 12, 28, 40, 52, 64, and 76). Patients who were on a stable dose of concomitant symptomatic medications (AChEI or memantine) were allowed to stay for the duration of the study.

Patients with mild to moderate AD who met entry criteria were randomized in a 1:1:1 ratio (500 per treatment arm) to 1 of the 3 treatment groups. Patients were randomized by site and severity of AD; mild or moderate AD based on the score from the Mini-Mental State Examination (MMSE) scale.

The primary hypothesis being tested is that the high dose of the experimental drug (HD) slows down the decline rate associated with AD as compared with placebo after 76 weeks of treatment. In other words, the decline for the experimental drug is smaller than the decline for placebo. This can be formulated as:

$$\begin{aligned} H_0 &: \mu_{76,\text{HD}} = \mu_{76,\text{Placebo}}, \\ H_1 &: \mu_{76,\text{HD}} < \mu_{76,\text{Placebo}}, \end{aligned}$$

where μ_{76} is the mean change in decline from baseline. Given that the expected

discontinuation rates for the high dose group were unknown, a contingency plan was included in the protocol. This plan stated that if a large proportion of patients in the high dose group dropped out of the study, then primary comparison would be between the low dose group and placebo. The specific criteria that would trigger the switch were pre-specified as follows: If (1) the discontinuation rate in the high dose group was greater than 50%; (2) the discontinuation rate in the high dose group exceeded the discontinuation rate in the placebo group by 20%; (3) the low dose group does not meet either of the criteria above (1) or (2).

In analyzing the treatment effect, a linear mixed-effects model is considered (Verbeke and Molenberghs 2000), allowing for a direct-likelihood approach to incomplete data, which is sometimes referred to as 'repeated measures mixed model' (MMRM), and is used for analysis of the primary endpoint. The general form is given in (9.2).

In our case, the model for the fixed effect includes 8 independent variables: baseline score, age at baseline, treatment, visit (post-baseline assessments; a categorical variable), treatment by visit interaction, MMSE stratification factor at baseline (mild or moderate), concomitant ACHEI or memantine use at baseline (yes or no), and investigator (site). In the final analysis, models with site as a random effect and a model without site are considered. Thus, all in all, we have four cases, referring to site as: (1) fixed effect; (2) random effect; (3) included while generating data but excluded in the data analysis; (4) excluded while generating data as well as in the data analysis.

10.3 Simulation Study

10.3.1 Generating Datasets

In line with the Alzheimer trial, our simulation starts from 176 sites. Each site is assumed to have and equal sample size of 9, that is, 3 patients per treatment arm. The total population sums up to $528 \times 3 = 1584$ patients. Every time, a complete dataset is generated. First, for each patient, the values of the covariates are generated from a number of practically plausible distributions. The baseline value for the co-primary outcome of cognition follows a $N(25, 9^2)$. The age of the patient at baseline is assumed to follow $N(70, 9^2)$, while the distribution for severity of the disease at baseline (mild=1/moderate=0) is Bernoulli(0.5). Finally, concomitant ACHEI or meantime use at baseline (Yes=1/No=0) follows a Bernoulli(0.75).

The mean and variance for age are chosen in line with knowledge about Alzheimer's

disease. Once the covariates are generated, both responses: efficacy (change from the baseline for one of the co-primary endpoints) and tolerance are jointly generated from a multivariate normal distribution. The underlying models for both response variables are given below. For efficacy, this is

$$\begin{aligned} Y_{ijk} &= \beta_1 B_{ij} + \beta_2 A_{ij} + \beta_3 M_{ij} + \beta_4 C_{ij} \\ &\quad + [\beta_5 I(k=1) + \beta_6 I(k=2) + \beta_7 I(k=3) + \beta_8 I(k=4) + \beta_9 I(k=5) + \\ &\quad \beta_{10} I(k=6)] \times I(T_{ij} = 140\text{mg}) \\ &\quad + [\beta_{11} I(k=1) + \beta_{12} I(k=2) + \beta_{13} I(k=3) + \beta_{14} I(k=4) + \beta_{15} I(k=5) + \\ &\quad \beta_{16} I(k=6)] \times I(T_{ij} = 100\text{mg}) \\ &\quad + [\beta_{17} I(k=1) + \beta_{18} I(k=2) + \beta_{19} I(k=3) + \beta_{20} I(k=4) + \beta_{21} I(k=5) \\ &\quad + \beta_{22} I(k=6)] \times I(T_{ij} = \text{Placebo}) \\ &\quad + b_i^Y + \varepsilon_{ijk}^Y. \end{aligned} \tag{10.1}$$

Here, Y_{ijk} is the efficacy response for patient j at site $i = 1, \ldots, 176$ and at visit $k = 1, \ldots, 6$. B_{ij}, A_{ij}, C_{ij}, M_{ij}, and T_{ij} are baseline, age, concomitant medication, MMSE, and treatment for patient j in site i, respectively. Further, $b_i^Y \sim N(0, \sigma^2)$ is a site-specific random effect, $\boldsymbol{\varepsilon}_{ij}^Y \sim N(0, \Sigma_{11})$ is a random error vector for the efficacy response, the patient specific random effect can be absorbed into it by choosing Σ_{11} an unstructured 6×6 covariance matrix. Similarly, the model for the tolerance response variable is:

$$\begin{aligned} Z_{ijk} &= [\alpha_1 I(k=1) + \alpha_2 I(k=2) + \alpha_3 I(k=3) + \alpha_4 I(k=4) + \alpha_5 I(k=5) + \\ &\quad \alpha_6 I(k=6)] \times I(T_{ij} = 140\text{mg}) \\ &\quad + [\alpha_7 I(k=1) + \alpha_8 I(k=2) + \alpha_9 I(k=3) + \alpha_{10} I(k=4) + \alpha_{11} I(k=5) + \\ &\quad \alpha_{12} I(k=6)] \times I(T_{ij} = 100\text{mg}) \\ &\quad + [\alpha_{13} I(k=1) + \alpha_{14} I(k=2) + \alpha_{15} I(k=3) + \alpha_{16} I(k=4) + \alpha_{17} I(k=5) + \\ &\quad \alpha_{18} I(k=6)] \times I(T_{ij} = \text{Placebo}) \\ &\quad + b_i^Z + \varepsilon_{ijk}^Z. \end{aligned} \tag{10.2}$$

Now, Z_{ijk} denotes the tolerance response for the j^{th} patient at the i^{th} site and at visit k, $\boldsymbol{\varepsilon}_{ij}^Z \sim N(0, \Sigma_{22})$, $\boldsymbol{\varepsilon}_{ij}^Z$ also is a random error vector for the tolerance response and here as well the patient-specific random effect is absorbed into it; Σ_{22} is an unstructured 6×6 covariance matrix. The tolerance response variable is assumed to

be related to treatment and visit. The joint distribution for both error terms is:

$$\begin{pmatrix} \varepsilon_{ij}^{Y} \\ \varepsilon_{ij}^{Z} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^{T} & \Sigma_{22} \end{pmatrix} \right],$$

with additionally Σ_{12} the 6×6 covariance structure between efficacy and tolerance. Two types of covariance structures were considered: (1) no correlation and (2) a correlation of 0.2 between efficacy and tolerance. For the latter, the covariance is assumed to have an autoregressive covariance structure in which the correlation between the measurements of efficacy and tolerance at the same visit is 0.2. The covariance matrices Σ_{11}, Σ_{22}, and Σ_{12} are given in Table 10.1. As far as our interest is in estimating the type I error rate, the data were generated under the null hypothesis of no difference in mean change between treatment arms at week 76 (visit 6), that is, $\beta_{10} = \beta_{16} = \beta_{22}$. The sets of parameters in Table 10.2 were used, and are chosen such that the measurements are within their appropriate ranges. A total of $S = 10,000$ datasets were simulated. Different sets of parameters (treatment means) are considered, to study the sensitivity of the error rate to the choices of the parameter values, and to make sure the results are robust to different settings. Four sets of parameter values (treatment means) are considered, leading to four sets of simulations. The second set of parameter values are set with more rapidly declining rates whereas the third set of parameter values are set with more slowly declining rates in efficacy over time (at visit 6), when compared to the first parameter values. In the fourth setting, parameter values are chosen in such a way that the efficacy of the patients is declining very slowly over time and the difference among the three treatment groups is very narrow.

In addition, one simulation (Simulation 5) is considered under different switching criteria. Although the main interest is in assessing the type I error rate associated with the switching criteria explained in Section 10.2, it is useful to consider other switching criteria for comparison purposes, where the switching condition is highly related to the test statistics for the significance of the treatment effect. In this respect, the following criterion was used to switch to compare low dose group and placebo: whether the efficacy of the patients in the low dose group is better than that of patients in the high dose group.

Table 10.1: *The covariance structure used for efficacy (Σ_{11}), tolerance (Σ_{22}), and between efficacy and tolerance, a correlation of 0, 2 and 0.8 (Σ_{12}^1, Σ_{12}^2, and Σ_{12}^3)*

$$\Sigma_{11} = \begin{pmatrix} 6 & 2.4 & 2.2 & 2.1 & 1.8 & 1.2 \\ 2.4 & 5 & 2.1 & 1.9 & 1.6 & 1.3 \\ 2.2 & 2.1 & 4.4 & 2.3 & 1.7 & 0.9 \\ 2.1 & 1.9 & 2.3 & 4.1 & 1.6 & 0.8 \\ 1.8 & 1.6 & 1.7 & 1.7 & 3.8 & 1.2 \\ 1.2 & 1.3 & 0.9 & 0.8 & 1.2 & 3.2 \end{pmatrix},$$

$$\Sigma_{22} = \begin{pmatrix} 1 & 0.4 & 0.32 & 0.31 & 0.28 & 0.21 \\ 0.4 & 1.2 & 0.6 & 0.42 & 0.38 & 0.24 \\ 0.32 & 0.6 & 1.3 & 0.5 & 0.44 & 0.32 \\ 0.31 & 0.42 & 0.5 & 1.1 & 0.56 & 0.38 \\ 0.28 & 0.38 & 0.44 & 0.56 & 1.2 & 0.42 \\ 0.21 & 0.24 & 0.32 & 0.38 & 0.42 & 1.4 \end{pmatrix},$$

$$\Sigma_{12}^1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\Sigma_{12}^2 = \begin{pmatrix} 0.49 & 0.107 & 0.022 & 0.0041 & 0.00086 & 0.00018 \\ 0.089 & 0.49 & 0.102 & 0.019 & 0.0039 & 0.00085 \\ 0.016 & 0.091 & 0.48 & 0.088 & 0.018 & 0.004 \\ 0.0032 & 0.018 & 0.092 & 0.42 & 0.089 & 0.019 \\ 0.00062 & 0.0034 & 0.018 & 0.082 & 0.43 & 0.092 \\ 0.00011 & 0.00068 & 0.0033 & 0.015 & 0.078 & 0.42 \end{pmatrix}$$

$$\Sigma_{12}^3 = \begin{pmatrix} 1.9370640 & 1.6952802 & 1.4024076 & 1.0250879 & 0.8569165 & 0.7342662 \\ 1.3847788 & 1.9091082 & 1.5695555 & 1.1424399 & 0.9439085 & 0.8086843 \\ 1.0140888 & 1.4055154 & 1.8308832 & 1.3327083 & 1.1029738 & 0.9365158 \\ 0.7686524 & 1.0671151 & 1.3872131 & 1.6068414 & 1.3250188 & 1.1279343 \\ 0.5832244 & 0.801930 & 1.052397 & 1.221365 & 1.619066 & 1.3825108 \\ 0.400769 & 0.5572728 & 0.733952 & 0.8530009 & 1.1404708 & 1.5724763 \end{pmatrix}.$$

Table 10.2: *Simulation study. Parameter values used for data generation.*

Sim.	Efficacy	tolerance
1,5	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$= 12, 16, 17,22, 25, 34	140 mg $(\alpha_1 - \alpha_6)$= 10, 12.2, 13.1, 14.5, 15, 15.5
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 18, 20, 24, 27, 34	100 mg $(\alpha_7 - \alpha_{12})$= 8, 10.2, 11.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 24, 27, 31, 34, 34	Placebo $(\alpha_{13} - \alpha_{18})$= 3.5, 4, 4.2, 4.5, 4.8, 4.9
2,6,7,8,9	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 18, 24, 30, 36	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 20, 26, 31, 36	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 25, 31, 36, 36	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2
3	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 17, 23, 27, 32	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 19, 24, 28, 32	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 24, 28, 34, 32	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2
4	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg:$(\beta_5 - \beta_{10})$=23, 23, 23.2, 23.4, 23.6, 24	140 mg $(\alpha_1 - \alpha_6)$=10, 12.2, 13.1, 14.5, 15, 15.5
	100 mg:$(\beta_{11} - \beta_{16})$=23, 23, 23.2, 23.5, 23.7, 24	100 mg $(\alpha_7 - \alpha_{12})$=8, 10.2, 11.5, 12.2, 13, 13.4
	Placebo:$(\beta_{17} - \beta_{22})$=23,23.2,23.2,23.6,23.7,24	Placebo $(\alpha_{13} - \alpha_{18})$=3.5, 4, 4.2, 4.5, 4.8, 4.9
8,9	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 0.4$	$\sigma_z = 0.4$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 18, 24, 30, 36	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 20, 26, 31, 36	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 25, 31, 36, 36	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2

10.3.2 Incorporating Incompleteness

On a generated complete dataset, a missingness mechanism was applied. First, this was done under a MAR mechanism, where missingness depends on the observed values only:

$$\begin{aligned}&\text{logit}[P(D_{ij} = k|D_{ij} > k-1, y_{ij,k-1}, T_{ij}, Z_{ij,k-1})]\\ &= \psi_0 + \psi_1 y_{ij,k-1} + \psi_2 I(T_{ij} = 140\text{mg}) + \psi_3 I(T_{ij} = 100\text{mg}) + \psi_4 Z_{ij,k-1}.\end{aligned}$$

Here, D_{ij} represents the time of drop out for patient j at site i, $y_{ij,k-1}$ is the previous observed measurement of this patient, representing the dependence of dropout on efficacy. Including treatment effect into the model helps to ensure that the switching condition is satisfied by assigning a different probability of dropping out for different treatment groups. $Z_{ij,k-1}$ is the previous/observed tolerance response, representing the dependence of dropout on tolerance.

In the first batch of simulations, three sets of parameter values (Scenarios 1, 2, and 3), leading to a probability of switching of about 10%, 25%, and 50%, respectively, are considered. Scenario 1 introduces dropouts at a rate of about 47% in the high dose (140 mg) group, 39% in the low dose (100 mg), and 24% in the placebo group. It leads to a probability of switching of about 10%. Similarly, Scenarios 2 and 3 introduce dropout profiles of about (48.5%, 40.5%, 25%) and (50%, 43.5%, 26%) in the high dose, low dose, and placebo groups, respectively. Whereas the first three scenarios result in switching with a given non-zero probability, two other sets of parameters (Scenarios 4 and 5), not resulting in switching, are also considered, for the sake of comparison. Scenario 4 introduces dropout of about 25% in the high dose group, 21% in the low dose group and 14% in the placebo group. In this case, dropout in the high dose group is not sufficiently large to meet the first requirement for switching. Scenario 5, on the other hand, introduces dropout of about 57% in the high dose group, 55% in the low dose group, and 48.5% in the placebo group. The percentage of dropout in all treatment groups is large but the percentage of dropout in the high dose group does not exceed that of placebo by 20% (second requirement for switching not satisfied).

To further see the impact of switching on the Type I error under a MNAR missing mechanism, such a mechanism was also considered. This is done by including the

current measurement, Y_{ijk}, into the logistic model:

$$\begin{aligned}&\text{logit}[P(D_{ij} = k|D_{ij} > k-1, y_{ij,k-1}, T_{ij}, Z_{ij,k-1})]\\ &= \psi_0 + \psi_1 y_{ij,k-1} + \psi_2 y_{ijk} + \psi_3 I(T_{ij} = 140\text{mg}) + \psi_4 I(T_{ij} = 100\text{mg}) + \psi_5 Z_{ij,k-1}\\ &\quad +\psi_5 Z_{ijk}.\end{aligned}$$

While it is evidently true that treatment is unknown to the trialist as well as to the patient, the patient does undergo the effects and therefore, in the true data generating model, a dependence on treatment is considered realistic, even though unavailable during the conduct of the trial, unless the trial is unblinded, or at least partially so (e.g., to the members of the monitoring committee).

Like the first five scenarios we considered under MAR, a corresponding collection of five scenarios under MNAR was also considered. Three sets of parameter values (Scenarios 6, 7, and 8), leading to probabilities of switching of about 10%, 25%, and 50% and two sets of parameters (Scenarios 9 and 10) that do not lead to switching. The percentage of dropout in each treatment arm for all scenarios is given in Table 10.3 and the ψ parameters are displayed in the Table 10.4. The final scenario corresponds to complete data, which is expected to produce a Type I error rate of 0.025. The latter is introduced as an internal checking device. These 11 scenarios are applied to the datasets generated under zero correlation between efficacy and tolerance. The same sets of parameters are also applied to data with a correlation of 0.2 between efficacy and tolerance, producing an additional 11 scenarios (Scenarios 12–22). An overview of the scenarios in the first batch of simulation is presented in Table 10.3.

The first set of simulations, which results from the first set of parameter values, is large, while the other simulation batches, three with different sets of efficacy parameters (treatment means) and one with different switching criteria, are done on a smaller scale, i.e., only a subset of the scenarios from the first simulation is considered. The second batch of simulations encompasses the first 11 scenarios. The third batch contains 6 scenarios (1–5 and 11). The fourth batch contains only 4 scenarios; two scenarios with switching and the remaining two without switching. The fifth batch, with alternate switching criterion, encompasses two scenario with switching (1 and 3) and another two without (4 and 5). To assess the sensitivity of the results to the choice of the number of sites as well as to the number of patients per site, two additional sets of simulations were conducted (simulation 6 and 7). These simulations are the same as the second batch of simulations, but with different numbers of sites and patients per site. In the first set of simulations, the number of sites is varied to

Table 10.3: *Overview of the scenarios considered for the first batch of simulations.*

					Av. % dropout		
Scenario	Corr.	Mech.	Switching	P(Switching)	high	low	placebo
1	0	MAR	Yes	0.10	47	39	24
2	0	MAR	Yes	0.25	48.5	40.5	25
3	0	MAR	Yes	0.50	50	43.5	26
4	0	MAR	No	-	25	21	14
5	0	MAR	No	-	57	55	48.5
6	0	MNAR	Yes	0.10	47	39	22.5
7	0	MNAR	Yes	0.25	48.5	40.5	23.5
8	0	MNAR	Yes	0.50	50	43.5	24.5
9	0	MNAR	No	-	25	21	12.5
10	0	MNAR	No	-	57	55	48.5
11	0	No	No	-	0	0	0
12	0.2	MAR	Yes	0.10	47	39	24
13	0.2	MAR	Yes	0.25	48.5	40.5	25
14	0.2	MAR	Yes	0.50	50	43.5	26
15	0.2	MAR	No	-	25	21	14
16	0.2	MAR	No	-	57	55	48.5
17	0.2	MNAR	Yes	0.10	47	39	22.5
18	0.2	MNAR	Yes	0.25	48.5	40.5	23.5
19	0.2	MNAR	Yes	0.50	50	43.5	24.5
20	0.2	MNAR	No	-	25	21	12.5
21	0.2	MNAR	No	-	57	55	48.5
22	0.2	No	No	-	0	0	0

50, 176, and 300, thereby keeping the number of patients per site at 3. In the second batch, the number of patients per site is varied to 3, 10, and 30, keeping the number of sites fixed at 176.

Moreover, two additional simulations (8 and 9) are conducted and add up to a total of 9 simulations. These are similar to the second batch of simulations but the treatment effect is excluded when introducing missingness into the data. Therefore, the dropouts are entirely induced either by efficacy, tolerance, or both. For MAR,

$$\text{logit}[P(D_{ij} = k | D_{ij} > k-1, y_{ij,k-1}, Z_{ij,k-1})] = \psi_0 + \psi_1 y_{ij,k-1} + \psi_2 Z_{ij,k-1}$$

. Assume that efficacy can be positively or negatively related with dropouts. That

Table 10.4: *Parameter values used for introducing missingness for Simulations 1 and 5.*

Scenario	Parameter values
1	ψ_0–ψ_4=-2.10, -0.02, 0.46, 0.3, 0.025
2	ψ_0–ψ_4=-2.05, -0.02, 0.45, 0.3, 0.025
3	ψ_0–ψ_4=-2.00, -0.02, 0.45, 0.35, 0.025
4	ψ_0–ψ_4=-2.75, -0.02, 0.30, 0.20, 0.025
5	ψ_0–ψ_4=-1.60, -0.01, 0.02, 0.01, 0.015
6	ψ_0–ψ_6=-2.20, -0.01, -0.01, 0.45, 0.30, 0.015, 0.02
7	ψ_0–ψ_6=-2.15, -0.01, -0.01, 0.45, 0.30, 0.015, 0.02
8	ψ_0–ψ_6=-2.10, -0.01, -0.01, 0.44, 0.35, 0.015, 0.02
9	ψ_0–ψ_6=-2.85, -0.01, -0.01, 0.30, 0.20, 0.015, 0.02
10	ψ_0–ψ_6=-1.60, -0.005, -0.005, 0.02, 0.01, 0.005, 0.01
11	-

is, patients may drop out due to poor efficacy at the previous visit ($\psi_1 > 0$) or due to temporary relief ($\psi_1 < 0$). We consider these cases, and some sub-cases, in turn. (1) The less efficacy there is, the more likely the patient drops out ($\psi_1 > 0$). In this case, the dropout has to be driven by tolerance. Otherwise, more missingness is likely in the placebo group and less in the high dose group and this implies that the switching condition cannot be satisfied. The tolerance response has no direct impact on the significance of the treatment effect but can have an indirect impact through the association with efficacy. (1.a) No correlation between efficacy and tolerance: Therefore, there is no association between significance of the treatment effect and switching criteria. As a result, no type I error rate inflation is expected. (1.b) Although not realistic, if we assume negative correlation between efficacy and tolerance, i.e., the higher tolerance response then the more efficacy (smaller value) there is, the switching criteria will have quite the opposite effect because we are switching to another treatment comparison when the chance of significance for the high dose is higher. Therefore, in that case, a lower type I error rate is expected. (1.c) Positive correlation between efficacy and tolerance, i.e., the higher tolerance response, the less efficacy (larger value) there is. This may result in inflation of the error rate. (2) The more efficacious the drug is, the more likely the patient drops out, stemming from temporary relief ($\psi_1 < 0$). In this case, if missingness is driven by efficacy, then the switching criteria does not systematically favor significance of the treatment effect. The switching criteria is associated with significance of the treatment effect but in a quite opposite way. Also here, a lower type I error rate is anticipated. The two addi-

Table 10.5: *Parameter values used for introducing missingness for Simulations 2, 3, 6 and 7.*

Scenario	Parameter values
1	ψ_0–ψ_4=-3.76, 0.015, 0.51, 0.36, 0.05
2	ψ_0–ψ_4=-3.72, 0.015, 0.51, 0.37, 0.05
3	ψ_0–ψ_4=-3.675, 0.015, 0.51, 0.36, 0.05
4	ψ_0–ψ_4=-4.45, 0.015, 0.4, 0.28, 0.05
5	ψ_0–ψ_4=-2.93, 0.015, 0.013, 0.011, 0.05
6	ψ_0–ψ_6=-4.10, 0.01, 0.01, 0.6, 0.41, 0.025, 0.025
7	ψ_0–ψ_6=-4.06, 0.01, 0.01, 0.6, 0.42, 0.025, 0.025
8	ψ_0–ψ_6=-4.01, 0.01, 0.01, 0.597, 0.42, 0.025, 0.025
9	ψ_0–ψ_6=-4.80, 0.01, 0.01, 0.45, 0.30, 0.025, 0.025
10	ψ_0–ψ_6=-3.20, 0.01, 0.01, 0.024, 0.018, 0.025, 0.025
11	-

Table 10.6: *Parameter values used for introducing missingness for Simulation 4.*

Scenario	Parameter values
1	$\psi_0 = -1.8$, $\psi_1 = -0.03$, $\psi_2 = 0.33$, $\psi_3 = 0.15$, $\psi_4 = 0.05$
2	$\psi_0 = -2.2$, $\psi_1 = -0.03$, $\psi_2 = 0.82$, $\psi_3 = 0.6$, $\psi_4 = 0.05$
3	$\psi_0 = -3.2$, $\psi_1 = -0.03$, $\psi_2 = 1.5$, $\psi_3 = 1.25$, $\psi_4 = 0.05$
4	$\psi_0 = -1.34$, $\psi_1 = -0.02$, $\psi_2 = 0.12$, $\psi_3 = 0.09$, $\psi_4 = 0.025$

tional simulations correspond to these two situations, i.e., the first one when dropout is driven by efficacy and the second one when dropout is induced by the tolerance response; a correlation between efficacy and tolerance of 0, 0.2, and 0.8 is considered.

10.3.3 Estimating the Type I Error Rate

The resulting datasets are analyzed using the aforementioned likelihood-based ignorable method, valid under MAR, thereby using all the available information without the need to either delete or impute measurements (Molenberghs and Verbeke 2005, Molenberghs and Kenward 2007). The same method was used to analyze the incomplete data resulting from MNAR. It enables us to see the impact of misspecification of the missing data mechanism. In addition to this likelihood-based ignorable method, last observation carried forward (LOCF) imputation was also used, limited to the

Table 10.7: *Parameter values used for introducing missingness for Simulation 8.*

Scenario	Parameter values
1	$\psi_0 = 1.86$, $\psi_1 = -0.12$, $\psi_2 = 0.02$
2	$\psi_0 = 1.86$, $\psi_1 = -0.118$, $\psi_2 = 0.02$
3	$\psi_0 = 1.86$, $\psi_1 = -0.117$, $\psi_2 = 0.02$
4	$\psi_0 = 1.2$, $\psi_1 = -0.125$, $\psi_2 = 0.02$
5	$\psi_0 = -1.2$, $\psi_1 = -0.02$, $\psi_2 = 0.02$

Table 10.8: *Parameter values used for introducing missingness for Simulation 9.*

Scenario	Parameter values
1	$\psi_0 = -4.47$, $\psi_1 = 0.015$, $\psi_2 = 0.14$
2	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.145$
3	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.148$
4	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.08$
5	$\psi_0 = -2.7$, $\psi_1 = 0.015$, $\psi_2 = 0.04$

case where site is considered a fixed effect. This allows us to assess the type I error rate under this more traditional analysis.

10.3.3.1 Scenarios with Switching

Those datasets satisfying the switching condition are analyzed based on the comparison of low dose and placebo (Model 1):

$$\begin{aligned} Y_{ijk} \quad = \quad & \beta_0 + \beta_1 I(T_{ij} = 100\text{mg}) + \beta_2 I(k = 1) + \beta_3 I(k = 2) + \beta_4 I(k = 3) \\ & + \beta_5 I(k = 4) + \beta_6 I(k = 5) + [\beta_7 I(k = 1) + \beta_8 I(k = 2) + \\ & \beta_9 I(k = 3) + \beta_{10} I(k = 4) + \beta_{11} I(k = 5)] I(T_{ij} = 100\text{mg}) \\ & + \beta_{12} A_{ij} + \beta_{13} M_{ij} + \beta_{14} C_{ij} + \beta_{15} B_{ij} + b_i + \varepsilon_{ijk}. \end{aligned} \tag{10.3}$$

The remainder are analyzed based on the comparison of high dose and placebo (Model 2):

$$\begin{aligned} Y_{ijk} &= \beta_0 + \beta_1 I(T_{ij} = 140\text{mg}) + \beta_2 I(k=1) + \beta_3 I(k=2) + \beta_4 I(k=3) \\ &\quad + \beta_5 I(k=4) + \beta_6 I(k=5) + [\beta_7 I(k=1) + \beta_8 I(k=2) + \\ &\quad \beta_9 I(k=3) + \beta_{10} I(k=4) + \beta_{11} I(k=5)] I(T_{ij} = 140\text{mg}) \\ &\quad + \beta_{12} A_{ij} + \beta_{13} M_{ij} + \beta_{14} C_{ij} + \beta_{15} B_{ij} + b_i + \varepsilon_{ijk}. \end{aligned} \tag{10.4}$$

In both models, visit 6 is the reference time, hence β_1 is the parameter of interest. For each dataset, we either reject or do not reject the null hypothesis $H_0 : \beta_1 = 0$. The overall Type I error rate is calculated as c/s, where c is the number of significant cases in both parts and s is the total number of simulations $(10,000)$.

10.3.3.2 Scenarios Without Switching

All datasets are analyzed based on the designed comparison of high dose and placebo. The type I error rate is calculated and compared with the type I error rate with switching. This enables us to see whether switching doses inflates the type I error rate and would need correction.

10.4 Simulation Results

Our primary goal was to compare the estimated Type I error rate with and without switching. A summary of the results of the first three batches, where site is considered as random and as fixed effect, is presented in Tables 10.9–10.11. The detailed results for all cases (site as fixed, random, and excluded) are presented in the Appendix D.1. When site is excluded from the final analysis, the results produced an error rate much lower than the nominal level, but the type I error rates for the remaining three are quite similar. They approximately amount to the allowed type I error rate of 0.025, with minimum of 0.02 and maximum of 0.0305. The reason for the low Type I error rate for case c, site excluded from analysis, is because in generating the data, site has been considered, while in the final analysis it is neither included as fixed nor as random effect. Thus, as a consequence, a certain amount of variability is left unexplained. The immediate consequence is an overestimation of the standard errors, which in turn results in underestimating the Type I error rate. When site was excluded completely, both in data generating and final analysis, the expected type I error rate was obtained.

In general, there was not much difference between the scenarios with and without switching, in the estimated type I error rate except for one scenario that will be discussed below. This finding appeared when it was applied under the same missing data mechanism and correlation between efficacy and tolerance. There was no systematic trend in error rate when the probability of switching increases from 10% to 50%. It is also of interest to see how the type I error rate would vary if we assume a different correlation between efficacy and tolerance. Noticeable differences in Type I error rate were not observed when a correlation of 0 and 0.2 between efficacy and tolerance was assumed in generating the data. If we consider the datasets generated under MAR and MNAR and analyzed using direct likelihood, a technique valid only under MAR, then, although results were generally very similar between both, we noted that the type I error rate under MNAR is slightly higher in most of the scenarios when compared to its counterpart under MAR in the second batch of simulation.

In contrast, and in line with expectation, the results obtained by analyzing these incomplete data using LOCF were very different (Table 10.9). The type I error rate was huge under all scenarios. The reason is that the efficacy for the patients in the high dose group is better, i.e., smaller change/decline in the primary endpoint, than the efficacy of the patients in the control and low dose groups in the first five visits. Also, the efficacy in the low dose group is better than in the control group. Although the data are generated under the null hypothesis, implying that the efficacy of the patient is the same in all treatment arms at the last visit, carrying the observation forward for about 50% of the missing observations will inflate significance of the treatment effect at the last visit. Two factors determine how much the type I error rate inflates: (1) the magnitude of difference in the efficacy between the treatment arms and placebo; and (2) the difference in the percentage of missingness between the treatment arms and placebo.

Let us turn to the fourth batch (Table 10.12). The error rate from the likelihood based analysis, like before, is controlled. The error rates from LOCF are still large. It is worth emphasizing that this large error resulted despite the data being generated under almost no treatment effect at any visit and very slight difference in the efficacy of the patients over time. This underscores just how much LOCF can inflate the error rate. Furthermore, the error rate is relatively lower when contrasted to their counterparts from the earlier results, with the same amount of missingness. This stems from the fact that the difference between the treatment arms and placebo is very small in this simulation. On the other hand, the difference in the percentage of missingness in the treatment arm forces the error rate to vary across the scenarios. For instance,

Table 10.9: *Summary of the estimated type I error rate from the first batch of results, for site as a fixed and a random effect.*

				Estimated Type I error rate		
					Site fixed	Site random
Scenario	Corr.	Mech.	Switching	Ign. lik.	LOCF	Ign. lik.
1	0	MAR	Yes	0.024	1.000	0.025
2	0	MAR	Yes	0.023	1.000	0.025
3	0	MAR	Yes	0.027	1.000	0.027
4	0	MAR	No	0.026	1.000	0.022
5	0	MAR	No	0.026	1.000	0.029
6	0	MNAR	Yes	0.025	1.000	0.024
7	0	MNAR	Yes	0.023	1.000	0.025
8	0	MNAR	Yes	0.025	1.000	0.025
9	0	MNAR	No	0.023	1.000	0.022
10	0	MNAR	No	0.026	1.000	0.027
11	0	No	No	0.021	0.021	0.026
12	0.2	MAR	Yes	0.029	1.000	0.028
13	0.2	MAR	Yes	0.028	1.000	0.024
14	0.2	MAR	Yes	0.029	1.000	0.025
15	0.2	MAR	No	0.024	1.000	0.023
16	0.2	MAR	No	0.025	1.000	0.022
17	0.2	MNAR	Yes	0.025	1.000	0.028
18	0.2	MNAR	Yes	0.023	1.000	0.022
19	0.2	MNAR	Yes	0.024	1.000	0.022
20	0.2	MNAR	No	0.025	1.000	0.022
21	0.2	MNAR	No	0.025	1.000	0.025
22	0.2	No	No	0.024	0.024	0.027

the error rate under Scenario 4 is three times that of Scenario 3. Considering the difference in the percentage of missingness between the treatment arms and placebo explains why this huge difference can happen. Carrying observations forward assumes that patients had no further deterioration. Thus, more observations in the treatment group have better efficacy than in the placebo group, leading to more spurious treatment effects. The first four simulation results and the last two simulation (6 and 7) assess the error rate inflation associated with switching dose group based on the criteria explained in Section 10.2. Results indicated that the error rates are controlled under the likelihood method. Next, let us see the other switching criteria. Results

Table 10.10: *Summary of the estimated type I error rate from the second batch of results, based on ignorable likelihood.*

				Estimated Type I error rate	
Scenario	Corr.	Mech.	Switching	Site fixed	Site random
1	0	MAR	Yes	0.0277	0.0264
2	0	MAR	Yes	0.0240	0.0251
3	0	MAR	Yes	0.0205	0.0249
4	0	MAR	No	0.0273	0.0250
5	0	MAR	No	0.0255	0.0250
6	0	MNAR	Yes	0.0274	0.0298
7	0	MNAR	Yes	0.0288	0.0251
8	0	MNAR	Yes	0.0220	0.0254
9	0	MNAR	No	0.0217	0.0267
10	0	MNAR	No	0.0271	0.0283
11	0	No	No	0.0234	0.0237

Table 10.11: *Summary of the estimated type I error rate from the third batch of results, based on ignorable likelihood.*

				Estimated type I error rate.	
Scenario	Corr.	Mech.	Switching	Site fixed	Site random
1	0	MAR	Yes	0.0248	0.0232
2	0	MAR	Yes	0.0204	0.0289
3	0	MAR	Yes	0.0288	0.0234
4	0	MAR	No	0.0295	0.0269
5	0	MAR	No	0.0259	0.0278
6	0	No	No	0.0260	0.0236

Table 10.12: *Summary of the type I error rate obtained from the fourth batch of simulations (with alternative parameter values); all scenarios are under MAR. (Av.per.drop.: average percentage of dropouts in the high, low and placebo response.*

			Estimated Type I error rate			
			Site fixed		Site random	
Scenario	Switching	Av.per.drop.	Likelihood	LOCF	Likelihood	LOCF
1	Yes	47,38,26	0.025	0.38	0.024	0.37
2	Yes	50,40,18	0.028	0.48	0.026	0.48
3	No	39.5,30,7	0.029	0.68	0.025	0.68
4	No	59,56,46.5	0.022	0.22	0.023	0.22

presented in Table 10.13 clearly indicate the inflation of the type I error rate, which is in line with the theoretical expectation that the type I error rate can inflate when the switching criterion is correlated with the test statistic for significance. If the two treatment arms are compared with placebo independently, an error rate of approximately 0.025 is expected. When switching in a manner unrelated to the significance of treatment effect, we still have an error rate of 0.025 because of the random nature of the switching act. In the worst case scenario, however, as the switching criteria is related with the level of significance for treatment effect, situations with a significant treatment effect will be included, leading to an error rate of about 0.05. This is true, provided an appropriate method for handling missing data is used; combined with inappropriate methods, it can even inflate further, and considerably so, as shown using LOCF, for instance. The sensitivity of the results to the choice of the number of

Table 10.13: *Summary of the type I error rate, obtained from the fifth batch of simulations with different switching criteria; all scenarios are under MAR and site is a fixed effect.*

Scenario	Switching	Type I error (likelihood)
1	Yes	0.03483
2	Yes	0.03320
3	No	0.02364
4	No	0.02899

sites, as well as to the number of patients per site was further assessed. The results, presented in Tables 10.14 and 10.15, show that it is insensitive to the number of sites and the number of patients per site. We do not see systematic differences in the Type I error rates between scenarios with and without switching, regardless the other

Table 10.14: *Summary of the estimated type I error rate for the sixth batch of results based on ignorable likelihood for different number of sites, where site is a random effect.*

				Type I error for # sites		
Scenario	Corr.	Mech.	Switching	50	176	300
1	0	MAR	Yes	0.028	0.025	0.025
2	0	MAR	Yes	0.025	0.021	0.026
3	0	MAR	Yes	0.029	0.027	0.025
4	0	MAR	No	0.027	0.022	0.022
5	0	MAR	No	0.026	0.023	0.027
6	0	MNAR	Yes	0.023	0.022	0.030
7	0	MNAR	Yes	0.029	0.028	0.031
8	0	MNAR	Yes	0.030	0.029	0.030
9	0	MNAR	No	0.029	0.022	0.031
10	0	MNAR	No	0.026	0.023	0.029
11	0	No	No	0.028	0.020	0.030

settings. The bias of the parameters of two covariates, treatment difference at the last visit and MMSE score, for the different scenarios of one particular simulation is presented in Table 10.16.

In the latter two simulations, the treatment effect is excluded when introducing missingness. Therefore, dropout is entirely induced either by efficacy or tolerance. In simulation 8 (Table 10.17), dropout is driven by efficacy such that the more efficacious the drug the higher the dropout rate due to temporary relief. The type I error rate is approximately the nominal rate, no error rate inflation associated with switching is noticed. In simulation 9 (Table 10.18), dropout is driven by the tolerance response. More dropout is associated with a higher tolerance response. In this simulation, a correlation of 0, 0.2, and 0.8 between the efficacy and tolerance is considered. As before, the results show that, for both correlations 0 and 0.2, no type I error shift was noted. However, for the specific case of a very high correlation of 0.8, the switching had some impact. This is because the switching criterion is highly related with efficacy through the tolerance response.

Table 10.15: *Summary of the estimated type I error rate from the seventh batch of results, based on ignorable likelihood, for different number of patients per site, where site is a random effect.*

				Type I error for # pat.		
Scenario	Corr.	Mech.	Switching	3	10	30
1	0	MAR	Yes	0.025	0.029	0.027
2	0	MAR	Yes	0.021	0.029	0.026
3	0	MAR	Yes	0.027	0.029	0.024
4	0	MAR	No	0.022	0.029	0.030
5	0	MAR	No	0.023	0.026	0.023
6	0	MNAR	Yes	0.022	0.031	0.027
7	0	MNAR	Yes	0.028	0.029	0.025
8	0	MNAR	Yes	0.029	0.030	0.031
9	0	MNAR	No	0.022	0.029	0.026
10	0	MNAR	No	0.023	0.026	0.028
11	0	No	No	0.020	0.025	0.026

Table 10.16: *Summary of the bias and standard error results for the seventh batch of results based on ignorable likelihood.*

				bias (s.e.)	
Scenario	Corr.	Mech.	Switching	Treat. diff. at V6	MMSE
1	0	MAR	Yes	-0.00429(0.1370)	0.00336(0.1004)
2	0	MAR	Yes	-0.00341(0.1372)	0.00372(0.1005)
3	0	MAR	Yes	-0.00431(0.1369)	0.00277(0.1005)
4	0	MAR	No	-0.00232(0.1221)	0.00242(0.0962)
5	0	MAR	No	-0.00708(0.1595)	0.00323(0.1059)
6	0	MNAR	Yes	-0.00878(0.1362)	0.00287(0.1000)
7	0	MNAR	Yes	-0.00777(0.1360)	0.00388(0.1000)
8	0	MNAR	Yes	-0.00806(0.1359)	0.00185(0.1001)
9	0	MNAR	No	-0.00381(0.1215)	0.00225(0.0960)
10	0	MNAR	No	-0.01016(0.1597)	0.00343(0.1057)
11	0	No	No	-0.00034(0.1102)	0.00266(0.0920)

Table 10.17: *Summary of the estimated type I error rate from the Eighth batch of results, based on ignorable likelihood.*

Scenario	Corr.	Mech.	Switching	Type I error rate
1	0	MAR	Yes	0.022
2	0	MAR	Yes	0.026
3	0	MAR	Yes	0.025
4	0	MAR	No	0.025
5	0	MAR	No	0.025

Table 10.18: *Summary of the estimated type I error rate from the ninth batch of results, based on ignorable likelihood.*

				Type I error rate		
Scenario	Mech.	P(Switching)	Switching	corr=0	corr=0.2	corr=0.8
1	MAR	Yes	0.1	0.019	0.024	0.032
2	MAR	Yes	0.25	0.023	0.025	0.039
3	MAR	Yes	0.5	0.028	0.025	0.038
4	MAR	No		0.019	0.025	0.028
5	MAR	No		0.024	0.021	0.031

10.5 Concluding Remarks

Adaptive studies that allow switching between dose groups are routinely used. In particular, a design that allows to choose a dose group to compare to placebo based on dropout rate can be considered. That is, the primary analysis may switch or shift to a different treatment contrast if the dropout rate is too high. It is a concern whether the type I error rate inflates with such design and requires a multiplicity adjustment. In this study, a simulation experiment was set up to assess the type I error rate inflation, inspired by an Alzheimer's disease trial associated with switching dose level. The type I error rate was estimated treating site as fixed effect, random effect, as well as by excluding it from analysis. All of this was done under different correlation levels between efficacy and tolerance, and under different missing mechanisms.

Based on the analysis using an ignorable likelihood method, the estimated type I error rate with and without switching was approximately the nominal error rate for the different scenarios except when dropout is strongly associated with efficacy and it was insensitive to the choice of parameters. Using LOCF imputation, the error rate was inflated, both with and without switching. However, no type I error rate inflation associated with switching was observed. Under a switching criterion, related with the test statistic for treatment effect, type I error rate inflation associated with switching is noticed.

We conclude that, although switching doses in a data-driven fashion at the final analysis, where the switching criteria is highly related with the primary endpoint, may in general lead to type I error rate inflation, the type I error rate inflation associated with switching was controlled for in most scenarios for the Alzheimer trial with longitudinal outcome where patients are expected to worsen over time. An exception occurred for the specific case where dropout is strongly associated with efficacy. Therefore the switching criteria used needs to be carefully studied regarding the possible association with significance of the treatment effect (Mallinckrodt *et al* 2003).

Our findings, when carefully consulted, can help minimize the impact of using untoward switch criteria.

Bibliography

Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. M.(2002). *Topics in Modelling of Clustered Data.* Chapman and Hall/CRC.

Agresti, A. (2002). *Categorical Data Analysis, second edition.* Hoboken: John Wiley and Sons.

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

Bahadur, R.R., (1961). *A representation of the joint distribution of responses to n dichotomous items. In: Studies in Item Analysis and Prediction, H. Solomon (Ed.). Stanford Mathematical Studies in the Social Sciences VI.* Stanford, CA: Stanford University Press.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57 (1)**, 289–300.

Blei, D. M. and Jordan, M. I. (2006). Variational Inference Dirichlet Process Mixtures. *Bayesian Analysis*, **1**, 121–144.

Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38–44.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Chang, M. (2008). *Adaptive Design Theory and Implementation using SAS and R.* Boca Raton: Chapman and Hall/CRC.

Chow, S.C. and Chang, M. (2008). Adaptive design methods in clinical trials—a review. *Orphanet Journal of Rare Diseases*, **3**, paper 11.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall/CRC.

Dale, J. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917

De Preter, V., Ghebretinsae, A. H., Abrahantes, J. C., Windey, K., Rutgeerts, P. and Verbeke, K. (2011). Impact of the Synbiotic combination Lactobacillus casei shirota and Oligofructose enriched Inulin on the Fecal Volatile Metabolite Profile in healthy subjects. *Molecular Nutrition and Food Reseach*, **55**, 5, Page 714–722.

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

Duran Pacheco, G., Hattendorf, J., Colford, J.M., Mausezahl, D., and Smith, T. (2009). Performance of analytical methods for overdispersed counts in cluster randomized trials. *Statistics in Medicine*, **28**, 2989–3011.

Edwards, A. W. F. (1971). Distances between populations on the basis of gene trequencres. *Biometrics*, **27**, 873–81.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge: Cambridge University Press.

Efendi, A., Molenberghs, G., and Verbeke, G. (2010). Pseudo-likelihood estimation for a combined model for repeated, overdispersed time-to-event data.

Ejchart, A. and Sadlej-Sosnowska, N. (2003). Statistical evaluation and comparison of comet assay results. *Mutation Research*, **534**, 85–92.

Faught, E., Wilder, B.J., Ramsay, R.E., Reife, R.A., Kramer, L.D., Pledger, G.W., and Karim, R.M. (1996). Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. *Neurology*, **46**, 1684–1690.

Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random effects approach. *Statistics in Medicine*, **23**, 3093–3104.

Fitzmaurice, G., Davidian, M. , Verbeke, G. and Molenberghs, G. (2009). *Longitudinal Data Analysis.* Chapman and Hall/CRC.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis* (2nd edn). Florida: Chapman and Hall/CRC.

Gelman, A. and Hill, J. (2006). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University.

Geyer,C. J. and Thompson, E. A. (1992). Constrained MonteCarlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B*, **54**, 657–699.

Ghebretinsae, A. H., Faes, C. , Molenberghs, G., De Boeck, M. and Geys, H. (2013). A Bayesian, Generalized Frailty Model for Comet Assays. *Journal of Biopharmaceutical Statistics*, **23**, 3, Page 618–636.

Ghebretinsae, A. H., Faes, C. , Molenberghs, G., Geys, H. and Van der Leede, B. (2012). Joint modeling of hierarchically clustered and overdispersed non-gaussian continuous outcomes for comet assay data. *Pharmaceutical Statistics*, **11**, 6, Page 449–455.

Ghebretinsae, A. H., Molenberghs, G., Dmitrienko, A., Offen, W. and Sethuraman, G. (2013). Assessment of Type I Error Rate Associated with Dose Group Switching in longitudinal Alzheimer Trial. *Journal of Biopharmaceutical Statistics*, accepted.

Hall, P., Ormerod, J.T. and Wand, M.P. (2011). Theory of Gaussian Variational Approximation for a Poisson Linear Mixed Model. *Statistica Sinica*, **21**, 369–389.

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

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

Jordan, M. I., and Xu, L. (1995). Convergence results for the EP approach to mixtures of experts architectures. *Neural Networks*, **8(9)**, 1409–1431.

Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997.

Kleinman, J. (1973). Proportions with extraneous variance: single and independent samples. *Journal of the American Statistical Association*, **68**, 46–54.

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.

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

Lee, Y. & Nelder, J.A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B*, **58**, 619-656.

Lee, Y., Nelder, J.A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood.* Boca Raton: Chapman & Hall/CRC.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the Americal Statistical Association*, **88**, 125–134.

Little, R.J.A. (1994a). A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data.* NewYork: John Wiley and Sons.

Liu, Q. & Pierce, D.A. (1994). A note on Gauss-Hermite quadrature. *Biometrika*, **81**, 624-629.

Lovell, D. P. and Omori, T. (2008). Statistical issues in the use of the comet assay. *Mutagenesis*, **23**, 171–182.

Mallinckrodt, C.H., Sanger, T.M., Dub, S., DeBrota, D.J., Molenberghs, G., Carroll, R.J., Potter, W.Z., and Tollefson, G.D. (2003). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, **53**, 754–760.

McGrory, C.A. and Titterington, D.M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics and Data Analysis*, **51**, 5352–5367.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* London: Chapman and Hall.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of American Statistical Association*, **89**, 633–644.

Molenberghs, G., Kenward, M.G. and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, **84**, 33–44.

Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley and Sons.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G., Verbeke, G., and Demétrio, C. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.

Molenberghs, G., Verbeke, G., Demétrio, C.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.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series B*, **135**, 370–384.

Ormerod, J. T. and Wand, M. P. (2009). Comment on paper by Rue, Martino and Chopin. J. *Journal of the Royal Statistical Society, Series B*, **71**, 377–378.

Ormerod, J.T. and Wand, M.P. (2010). Explaining Variation Approximations. *The American Statistician*, **64**, 140–153.

Ormerod, J.T. and Wand, M.P. (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, **21**, 2–17.

Ostling, O. and Johanson, KJ. (1984). Microelectrophoretic study of radiationinduced DNA damages in individual mammalian cells. *Biochem. Biophys. Res. Commun.*, **123**, 291–298.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.

Price, C.J., Kimmel, C.A., Tyl, R.W., and Marr, M.C. (1985). The Developmental Toxicity of Ethylene Glycol in Rats and Mice. *Toxicology and Applied Pharmacology*, **81**, 113–127.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, D.B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. In: *Imputation and Editing of Faulty or Missing Survey Data*. Washington, DC: U.S. Department of Commerce, 1–23.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Singh, N.P., McCoy, M.T., Tice, R.R., and Schneider, E.L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp. Cell Res.*, **175**, 184–191.

Skellam, J.G. (1948). A probability distribution derived from the binomial distributionby regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society, Series B*, **10**, 257–261.

Smith, C.C., Adkins, D.J., Martin, E.A., and O'Donovan, M.R. (2008). Recommendations for the design of the rat comet assay. *Mutagenesis*, 1–8.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64(4)**: 583–639.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, **12(3)**, 1–16.

Tanner, M. A. (1991). *Tools for Statistical Inference*. New York: Springer-Verlag.

Tsiatis, A. A. and Davidian, M.(2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, **14**, 809–834.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Wang, B. and Titterington, D.M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Analysis*, **1**, 625-650.

Wang, M., Wu, Y.C., and Tsai, G.F. (2008). A regulatory view of adaptive trial design. *Journal of the Formos Medical Association*, **107**, 3–8.

Wiklund, S.J. and Agurell, E. (2003). Aspects of design and statistical analysis in the Comet assay. *Mutagenesis*, **18**, 167–175.

Zeger, S.L., and Karim, M.R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79–102.

Appendices

Summary

In this thesis, modeling issues and design aspects that arise in toxicologic studies and clinical trials are addressed. The text is structured in two parts.

The first part of the thesis is motivated by a toxicologic study measuring DNA damage: the so-called comet assay. It was developed by Ostling and Johansson in 1984 and later modified by Singh *et al* in 1988. It has since increased in popularity for the evaluation of DNA damage and genotoxicity testing, and is now used as a standard tool in the pharmaceutical industry for the assessment of the safety of potential new drugs. In the comet assay allows DNA migration is visualised (typical comet-like structures), allowing the quantification of DNA damage at the single cell level. Three measures are commonly used: the tail migration (i.e., tail length), percentage tail intensity, and tail moment. Different issues complicate the analysis. (1) The design of the study is hierarchical in the sense that animals are nested within doses, a number of slides per animals are used and several cells are measured per slide. (2) Comet measures from an animal are not normally distributed but are rather asymmetric, skewed, bi- or multimodal, a mixture of different distributions, etc. Traditional analyses typically completely or partly ignore these issues and summarize measurements within a cluster. In this part, we have explored statistical models for hierarchically structured and overdispersed outcomes, such as for the comet assay data. Both univariate and multivariate methods are considered. Because of the computational complexity of the proposed models, an alternative estimation technique is explored.

The standard approach of modeling non-normal data is through a generalized linear model. The best-known examples include linear regression, logistic regression, and Poisson regression. An important extension of these models is the generalized linear mixed model, by the inclusion of normally distributed random effects, accounting for the multi-level structure in the data. A common issue with non-Gaussian data is overdispersion in the sense that the variability in the data is not well described by

the distributional mean-variance relationship. This can happen both in the univariate and in the multi-level setting. One approach to account for overdispersion in a univariate generalized linear model is by the use of a conjugate random effect, such as in for example the negative binomial and beta-binomial model. In a recent publication, Molenberghs, Verbeke, and Demétrio (2007) proposed a similar approach to account for overdispersion in a multilevel setting, by the use of two random effects, a normally distributed random effect to accommodate for the hierarchy and part of the overdispersion, and a conjugate random effect to account for the remaining overdispersion in the data. This is often referred to as the *combined model*.

In Chapter 4, a flexible modeling approach for hierarchically clustered and overdispersed non-Gaussian outcomes for comet assays is proposed. The model exhibits a full hierarchical structure, an appropriate distribution, and possible overdispersion using a combined model. In this approach, while a conjugate gamma random effect is used for the overdispersion random effect, both gamma and normal random effects are considered for the hierarchical random effect. The more conventional models with either the overdispersion, or just one hierarchical random effect are considered as sub-models. It provides a wide choice of models to select from. The use of more elaborate model with overdispersion and hierarchical structure improved the fit for one response.

Up to this point, the outcomes are modeled univariately. However, the comet assay data exhibit a multivariate structure. In general, multivariate longitudinal or clustered data are commonly encountered in clinical trials and toxicologic studies. Typically, there is no single standard endpoint to assess the toxicity or efficacy of the compound of interest, but co-primary endpoints are available to assess the toxic effects or the working of the compound. Modeling the responses jointly is thus appealing to draw overall inferences using all responses and to capture the association among the responses. A further extension to a multivariate setting with hierarchically clustered and overdispersed non-Gaussian outcomes is proposed in Chapter 5. The two outcomes are jointly analyzed by assuming that the normal random effects for both endpoints are correlated. The association structure between the response is analytically derived.

In Chapter 6, alternatives for the overdispersion distribution are considered: (1) A discrete distributions that eventually lead to finite mixture models as an alternative way of accounting for overdispersion. (2) Zero-inflated models are also explored to account for the excess zeros. (3) A mixture of conjugate distributions, where deviation from a single conjugate distribution are allowed for, while the property of conjugacy

can still be employed to ease computations. This further improves the flexibility of the combined model.

Most of the statistical models considered are using the combined model of Molenberghs *et al* (2010) or extensions thereof. In general, the main difficulty with this kind of models is the computationally complex estimation due to the intractable multivariate integrals, as is the case for generalized linear mixed models that involve such integrals with no analytic expression. In Chapter 7, the use of Gaussian variational approximation is explored as an alternative estimation technique. It approximates the integrand by introducing a set of variational densities in such a way that their evaluation is tractable. It is quite useful in the cases where there is computational difficulty using standard estimation or when computational time would be excessive otherwise.

The second part is related to incomplete data in clinical trials. In clinical trials, although the primary focus of many trials is on a specific time of measurement, usually the last, the outcome of interest is recorded in a longitudinal fashion, and missingness in general and dropout in particular is a common feature. Such missing data, in general, have a potential to affect/distort inferences drawn. It can lead to biased results and there can be a severe loss of power if the proportion of incompleteness is high. Some trials allow for data-driven adaptation when the dropout rate is high. For example, if the dropout rate in one of the treatment arms is very high, that treatment group is dropped and another treatment arm with less missing data used instead. That is, one allows for switching treatment comparison at the end of the study. However, reliability of the inference drawn is questioned as it can have an impact on the type I error rate. Chapter 10 focuses on the impact of such data driven adaptations. In particular, the type I error rate associated with dose group switching is assessed when the primary analysis is in terms of a longitudinal outcome. It is believed by some that switching dose groups may lead to an inflated Type I error rate and thus the significance level needs to be adjusted. On the other hand, it may not be applicable to some trials, including the Alzheimer's disease trial considered here. The Type I error rate can be lower for such trials, thus not inflating the overall Type I error rate, when the primary analysis is modified by switching doses based on high dropout rate in the high dose arm. In this study, a simulation experiment was set up to assess the type I error rate inflation, inspired by an Alzheimer's disease trial associated with switching dose level. The type I error rate was assessed under a number of scenarios, in terms of differing correlations between efficacy and tolerance, different missingness

mechanisms, and different probabilities of switching. A collection of parameter values was used to assess sensitivity of the analysis. We conclude that, although switching doses in a data-driven fashion at the final analysis, where the switching criteria is highly related with the primary endpoint, may in general lead to type I error rate inflation, the type I error rate inflation associated with switching was controlled for in most scenarios for the Alzheimer trial with longitudinal outcome where patients are expected to worsen over time. An exception occurred for the specific case where dropout is strongly associated with efficacy. Therefore, the switching criteria used needs to be carefully studied regarding the possible association with significance of the treatment effect.

Appendix A

Supplementary Materials of Chapter 4

In this appendix, we present the code used in Chapter 4.

```
#Model 8 - Weibull gamma(re2) #
model {
# N observations
  for (i in 1:N) {
  intensity[i]$\sim$dweib(r,lamda[i])
  lamda[i] <- h[slide[i]]*exp(eta[i])
  eta[i] <- beta0+beta1*low[i] + beta2*med[i]+
            beta3*high[i]+beta4*pos[i] }

# P Slides
    for (k in 1:P) { h[k] $\sim$ dgamma(alpha, alpha) }
# priors
   beta0$\sim$dnorm(0.0, 1.0E-6)
   beta1$\sim$dnorm(0.0, 1.0E-6)
   beta2$\sim$dnorm(0.0, 1.0E-6)
   beta3$\sim$dnorm(0.0, 1.0E-6)
   beta4$\sim$dnorm(0.0, 1.0E-6)
   alpha$\sim$dgamma(0.1,0.1)
   r$\sim$dexp(0.01)}
```

```
# Model 16 -  Weibull gamma(OD) normal(RE1) gamma(RE2) model #
model {
# N observations
  for (i in 1:N) {
  intensity[i]$\sim$dweib(r,lamda[i])
  lamda[i] <- h[slide[i]]*theta[i]*exp(eta[i])
  theta[i]$\sim$dgamma(alpha1, alpha1)
  eta[i] <- beta0+beta1*low[i]+beta2*med[i]+
            beta3*high[i]+beta4*pos[i]+u[rat[i]] }
# M rats
   for (j in 1:M) { u[j] $\sim$ dnorm(0,tau) }
# P slides
   for (k  in 1:P) { h[k] $\sim$ dgamma(alpha2, alpha2) }
# priors
   beta0$\sim$dnorm(0.0, 1.0E-6)
   beta1$\sim$dnorm(0.0, 1.0E-6)
   beta2$\sim$dnorm(0.0, 1.0E-6)
   beta3$\sim$dnorm(0.0, 1.0E-6)
   beta4$\sim$dnorm(0.0, 1.0E-6)
   tau$\sim$dgamma(0.01, 0.01)
   alpha1$\sim$dgamma(2, 2)
   alpha2$\sim$dgamma(0.1,0.1)
   r$\sim$dexp(0.01)}
```

Appendix B

Supplementary Materials of Chapter 5

In this appendix, we present the code used, derivation of the formulas and additional tables for the simulation studies discussed in Chapter 5.

B.1 Software Code

```
proc nlmixed data=comet2r maxiter=5000 qpoints=30;
/*specification of initial values*/
parms int0=-28  int125=-40  int25=-40.2 int5=-41 int200=-36
in0=-2.9 in125=-5.5 in25=-6.2 in5=-6.5 in200=-4.9 logrho1=1.4
logrho2=2 logalpha1=1 sigma1=0.6 sigma2=0.117 r=0.2;

/*specification of the linear part of the model for*/
           /*the first response*/
if (dose=0) then eta1=int0+b;
else if (dose=1.25) then eta1=int125+b;
else if (dose=2.5) then eta1=int25+b;
else if (dose=5) then eta1=int5+b;
else if (dose=200) then eta1=int200+b;

/*specification of the linear part of the model for*/
```

```
        /*the second response*/
if (dose=0) then eta2=in0+c;
else if (dose=1.25) then eta2=in125+c;
else if (dose=2.5) then eta2=in25+c;
else if (dose=5) then eta2=in5+c;
else if (dose=200) then eta2=in200+c;

/*use appropriate link functions*/
k1=exp(eta1);
k2=exp(eta2);
rho1=exp(logrho1);
rho2=exp(logrho2);
alpha1=exp(logalpha1);

/*likelihood specification */
if res=1 then
loglik=logrho1+log(response)*(rho1-1)+(alpha1+1)*logalpha1
+ eta1-(alpha1+1)*log(alpha1+(response**(rho1))*k1);
else if
res=2 then loglik=logrho2+log(response)*(rho2-1)+eta2-
((response**rho2)*k2);
model response~ general(loglik);

/*normal random effect specification*/
random b c ~normal([0,0],[sigma1**2,r*sigma1*sigma2,sigma2**2])
subject=slide_id;
/*some estimate statements of interest for first response*/
estimate 'Veh vs Low' int125 - int0;
estimate 'Veh vs Med' int25 - int0;
estimate 'Veh vs High' int5 - int0;
estimate 'Veh vs P. Control' int200-int0;
estimate 'low vs Med' int25 - int125;
estimate 'low vs High' int5- int125;
estimate 'Med vs high' int5-int25;

/*some estimate statements of interest for second response*/
```

```
estimate 'Veh vs Low' in125 - in0;
estimate 'Veh vs Med' in25 - in0;
estimate 'Veh vs High' in5 - in0;
estimate 'Veh vs P. Control' in200-in0;
estimate 'low vs Med' in25 - in125;
estimate 'low vs High' in5- in125;
estimate 'Med vs high' in5-in25;

/*Intraclass correlation and correlation between two responses*/
            /*estimate statements*/
betta1=beta(alpha1-(2/rho1),(2/rho1));
betta2=beta(alpha1-(1/rho1),(1/rho1));
ex1=exp((sigma1/rho1)**2);
estimate 'ICC1'   (ex1-1)/((2*rho1*betta1*ex1/(betta2**2))-1);
gamma1=gamma(2/rho2);
gamma2=gamma(1/rho2);
ex2=exp((sigma2/rho2)**2);
estimate 'ICC2' (ex2-1)/((2*rho2*gamma1*ex2/(gamma2**2))-1);
ex3=exp((r*sigma1*sigma2)/(rho1*rho2));
estimate 'Corr' (ex3-1)/sqrt(((2*rho1*betta1*ex1/(betta2**2))-1)
*((2*rho2*gamma1*ex2/(gamma2**2))-1));
run;
```

B.2 Derivation of the Correlation Between Both Endpoints

Let Y_{1ij} and Y_{2ij} be the j^{th} measurements of subject i for outcome 1 and 2 respectively. The linear part for the two responses are:

$$\begin{aligned} \eta_{1ij} &= \boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + b_{1i}, \\ \eta_{2ij} &= \boldsymbol{x}'_{2ij}\boldsymbol{\xi}_2 + b_{2i}, \end{aligned}$$

with

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_1^2 & rd_1d_2 \\ rd_1d_2 & d_2^2 \end{pmatrix}\right].$$

The correlation between the two endpoints is, by definition:

$$\text{Corr}(Y_{1ij}, Y_{2ij}) = \frac{\text{Cov}(Y_{1ij}, Y_{2ij})}{\sqrt{\text{Var}(Y_{1ij})}\sqrt{\text{Var}(Y_{2ij})}}. \quad \text{(B.1)}$$

We will re-write this as

$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{T1}{\sqrt{T2}\sqrt{T3}}. \quad \text{(B.2)}$$

It is also known that

$$\text{Cov}(Y_{ij}, Y_{ik}) = \text{E}\left(\text{Cov}(Y_{ij}, Y_{ik}|b_{1i}, b_{2i})\right) + \text{Cov}\left[\text{E}(Y_{ij}|b_{1i}, b_{2i}), \text{E}(Y_{ik}|b_{1i}, b_{2i})\right],$$

which we denote as $T1 = L1 + L2$.

Given the random effect, the two measurements are independent. Therefore

$$\text{Cov}(Y_{ij}, Y_{ik}|b_{1i}, b_{2i}) = 0.$$

By integrating the gamma random effect, we have:

$$f(y_{1ij}|b_{1i}) = \frac{\lambda_1 \rho_1 y_{1ij}^{\rho_1 - 1} e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + b_{1i}} \alpha_1^{\alpha_1 + 1}}{(\alpha_1 + \lambda_1 y_{1ij}^{\rho_1} e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + b_{1i}})^{\alpha_1 + 1}}, \quad \text{(B.3)}$$

with similar formula for $f(y_{2ij}|b_{2i})$.

Further, we have,

$$f(y_{1ij}|b_{1i}, b_{2i}) = f(y_{1ij}|b_{1i}), \quad \text{(B.4)}$$

$$f(y_{2ij}|b_{1i}, b_{2i}) = f(y_{2ij}|b_{2i}). \quad \text{(B.5)}$$

Therefore, the conditional expectation is given by

$$\text{E}(Y_{1ij}|b_{1i}) = \int y_{1ij} f(y_{1ij}) dy_{1ij} = \frac{\alpha_1^{\frac{1}{\rho_1}} B(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1})}{\rho_1 (\lambda_1 e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + b_{1i}})^{\frac{1}{\rho_1}}}, \quad \text{(B.6)}$$

with a similar formula for $\text{E}(Y_{2ij}|b_{2i})$.

The covariance between $\mathrm{E}(Y_{1ij}|b_{1i})$ and $\mathrm{E}(Y_{2ij}|b_{2i})$ is

$$\mathrm{Cov}[\mathrm{E}(Y_{1ij}|b_{1i}),\mathrm{E}(Y_{2ij}|b_{2i})] = \frac{\alpha_1^{\frac{1}{\rho_1}}\alpha_2^{\frac{1}{\rho_2}}B(\alpha_1-\frac{1}{\rho_1},\frac{1}{\rho_1})B(\alpha_2-\frac{1}{\rho_2},\frac{1}{\rho_2})}{\rho_1(\lambda_1 e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1})^{\frac{1}{\rho_1}}\rho_2(\lambda_2 e^{\boldsymbol{x}'_{2ij}\boldsymbol{\xi}_2})^{\frac{1}{\rho_2}}}\mathrm{Cov}(e^{-\frac{b_{1i}}{\rho_1}},e^{-\frac{b_{2i}}{\rho_2}}).$$

As a result:

$$T1 = L2 = \frac{\alpha_1^{\frac{1}{\rho_1}}\alpha_2^{\frac{1}{\rho_2}}B(\alpha_1-\frac{1}{\rho_1},\frac{1}{\rho_1})B(\alpha_2-\frac{1}{\rho_2},\frac{1}{\rho_2})}{\rho_1(\lambda_1 e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1})^{\frac{1}{\rho_1}}\rho_2(\lambda_2 e^{\boldsymbol{x}'_{2ij}\boldsymbol{\xi}_2})^{\frac{1}{\rho_2}}}e^{\frac{1}{2}(\frac{d_1^2}{\rho_1^2}+\frac{d_2^2}{\rho_2^2})}[e^{\frac{rd_1d_2}{\rho_1\rho_2}}-1], \quad \text{(B.7)}$$

$$T2 = \mathrm{Var}(Y_{1ij}) = \frac{\alpha_1^{\frac{2}{\rho_1}}e^{\frac{d_1^2}{\rho_1^2}}\left[B\left(\alpha_1-\frac{2}{\rho_1},\frac{2}{\rho_1}\right)e^{\frac{d_1^2}{\rho_1^2}}-B\left(\alpha_1-\frac{1}{\rho_1},\frac{1}{\rho_1}\right)^2\right]}{\rho_1\left(\lambda_1 e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1}\right)^{\frac{1}{\rho_1}}}, \quad \text{(B.8)}$$

and

$$T3 = \mathrm{Var}(Y_{2ij}) = \frac{\alpha_2^{\frac{2}{\rho_2}}e^{\frac{d_2^2}{\rho_2^2}}\left[B\left(\alpha_2-\frac{2}{\rho_2},\frac{2}{\rho_2}\right)e^{\frac{d_2^2}{\rho_2^2}}-B\left(\alpha_2-\frac{1}{\rho_2},\frac{1}{\rho_2}\right)^2\right]}{\rho_2\left(\lambda_2 e^{\boldsymbol{x}'_{2ij}\boldsymbol{\xi}_2}\right)^{\frac{1}{\rho_2}}}. \quad \text{(B.9)}$$

Substituting $T1$, $T2$ and $T3$ in (B.2) gives a formula for the correlation between the two endpoints in (5.2).

Derivation of the Intraclass Correlation

The correlation between the j^{th} and k^{th} measurements of subject i for outcome 1, Y_{1ij} and Y_{1ik} is

$$\mathrm{Corr}(Y_{1ij},Y_{1ik}) = \frac{\mathrm{Cov}(Y_{1ij},Y_{1ik})}{\sqrt{\mathrm{Var}(Y_{1ij})}\sqrt{\mathrm{Var}(Y_{1ik})}} \quad \text{(B.10)}$$

with

$$\mathrm{Cov}(Y_{ij},Y_{ik}) = \mathrm{E}(\mathrm{Cov}(Y_{1ij},Y_{1ik}|b_{1i})) + \mathrm{CovE}(Y_{1ij}|b_{1i}),\mathrm{E}(Y_{1ik}|b_{1i}) \quad \text{(B.11)}$$

Given that the random effect between the measurements are independent, $\mathrm{E}(\mathrm{Cov}(Y_{1ij}, Y_{1ik}|b_{1i})) = 0$ and

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) = \frac{\alpha_1^{\frac{2}{\rho_1}} B(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1})^2}{{\rho_1}^2 \left(\lambda_1^{\frac{2}{\rho_1}} e^{\boldsymbol{x}'_{1ij}\boldsymbol{\xi}_1 + \boldsymbol{x}'_{1ik}\boldsymbol{\xi}_1}\right)^{\frac{1}{\rho_1}}} e^{\frac{d_1^2}{\rho_1^2}} \left[e^{\frac{d_1^2}{\rho_1^2}} - 1 \right] \tag{B.12}$$

$\mathrm{Var}(Y_{1ij})$ is just the variance given in (B.9). Solving this leads to the intraclass correlation

$$\mathrm{Corr}(Y_{1ij}, Y_{1ik}) = \frac{\left[e^{\frac{d}{\rho_1^2}} - 1 \right]}{\left[\frac{2\rho_1 B(\alpha_1 - \frac{2}{\rho_1}, \frac{2}{\rho_1}) e^{\frac{d}{\rho_1^2}}}{B(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1})^2} - 1 \right]} \tag{B.13}$$

B.3 Simulation Results

Table B.1: *Type I error rate, power of the test and parameter estimate. Correlation between $r = 0.9$ and $\alpha = 1.5$. W: Weibull; G: gamma; N: normal; J: joint.*

	Response	Parameter	Traditional	W	WN	WGN	JWGN
Power	1		0.34	0.43	0.23	0.33	0.262
	2		0.44	0.725	0.61	0.61	0.651
	Combined						0.648
Estimate(s.e.)	1	β_{10}		-1.0258(0.0939)	-1.076(0.1185)	-1.0182(0.1367)	-1.0216(0.1296)
		β_{11}		-1.2015(0.0977)	-1.2649(0.1229)	-1.3044(0.1395)	-1.2946(0.1312)
		$\beta_{11} - \beta_{10}$		-0.1757(0.1061)	-0.1889(0.1436)	-0.2862(0.1774)	-0.2654(0.1794)
	2	β_{20}		-0.9892(0.0943)	-1.0027(0.1049)	-1.0027(0.1049)	-1.0049(0.1008)
		β_{21}		-1.2850(0.1026)	-1.3014(0.1127)	-1.3014(0.1127)	-1.3080(0.1065)
		$\beta_{21} - \beta_{20}$		-0.2958(0.1063)	-0.2987(0.1236)	-0.2987(0.1236)	-0.3120(0.1236)
Type I	1		0.06	0.17	0.045	0.04	0.0394
	2		0.06	0.125	0.07	0.07	0.054
	Combined						0.0259
Estimate(s.e.)	1	β_{10}		-1.0235(0.0939)	-1.0731(0.1184)	-1.0141(0.1367)	-1.022(0.1341)
		β_{11}		-1.0106(0.0933)	-1.0609(0.1181)	-1.0052(0.1359)	-1.0138(0.1333)
		$\beta_{11} - \beta_{10}$		0.0129(0.1059)	0.0121(0.1432)	0.0089(0.1757)	0.0047(0.1771)
	2	β_{20}		-0.9886(0.0943)	-1.0024(0.1054)	-1.0024(0.1054)	-1.0096(0.1052)
		β_{21}		-0.9883(0.0946)	-0.9997(0.1055)	-0.9997(0.1055)	-1.0146(0.1034)
		$\beta_{21} - \beta_{20}$		0.00039(0.1056)	0.00276(0.1238)	0.00276(0.1238)	-0.0035(0.1227)

Table B.2: *Type I error rate, power of the test and parameter estimate. Correlation between $r = 0.6$ and $\alpha = 1.5$. W: Weibull; G: gamma; N: normal; J: joint.*

	Response	Parameter	Traditional	W	WN	WGN	JWGN
Power	1		0.305	0.435	0.23	0.335	0.276
	2		0.46	0.73	0.64	0.64	0.6
	Combined						0.612
Estimate(s.e.)	1	β_{10}		-1.0248(0.0939)	-1.0754(0.1186)	-1.0170(0.1366)	-1.0359(0.1368)
		β_{11}		-1.2044(0.0979)	-1.2667(0.1230)	-1.3052(0.1393)	-1.3176(0.1387)
		$\beta_{11} - \beta_{10}$		-0.1795(0.1061)	-0.1913(0.1437)	-0.2882(0.1773)	-0.2746(0.1829)
	2	β_{20}		-0.9983(0.0946)	-1.0099(0.10417)	-1.0099(0.10417)	-1.0024(0.1022)
		β_{21}		-1.2803(0.10238)	-1.2961(0.1118)	-1.2961(0.1118)	-1.2923(0.1111)
		$\beta_{21} - \beta_{20}$		-0.2820(0.1063)	-0.2862(0.1220)	-0.2862(0.1220)	-0.2857(0.1280)
Type I	1		0.04	0.165	0.035	0.035	0.041
	2		0.06	0.095	0.07	0.07	0.062
	Combined						0.0396
Estimate(s.e.)	1	β_{20}		-1.0241(0.0938)	-1.0726(0.1181)	-1.0133(0.1368)	-1.0194(0.1282)
		β_{11}		-1.008(0.0933)	-1.0572(0.1177)	-0.9988(0.1361)	-0.9893(0.1266)
		$\beta_{11} - \beta_{10}$		0.0161(0.1059)	0.0154(0.1427)	0.0144(0.1759)	0.0307(0.1777)
	2	β_{20}		-0.9835(0.0942)	-0.9961(0.1045)	-0.9961(0.1045)	-1.0072(0.1031)
		β_{21}		-0.9823(0.0945)	-0.9929(0.1046)	-0.9929(0.1046)	-1.0011(0.0989)
		$\beta_{21} - \beta_{20}$		0.0012(0.1056)	0.0032(0.1225)	0.0032(0.1225)	0.0115(0.1223)

Table B.3: *Type I error rate, power of the test and parameter estimate. Correlation between $r = 0.3$ and $\alpha = 1.5$. W: Weibull; G: gamma; N: normal; J: joint.*

	Response	Parameter	Traditional	W	WN	WGN	JWGN
Power	1		0.305	0.405	0.24	0.305	0.290
	2		0.445	0.75	0.65	0.65	0.667
	Combined						0.642
Estimate(s.e.)	1	β_{20}		-1.0314(0.0941)	-1.0789(0.1177)	-1.0223(0.1360)	-1.0124(0.1340)
		β_{11}		-1.2070(0.0979)	-1.2675(0.122)	-1.3077(0.1386)	-1.3005(0.1373)
		$\beta_{11} - \beta_{10}$		-0.1756(0.1061)	-0.1886(0.1422)	-0.2853(0.1761)	-0.2818(0.1775)
	2	β_{20}		-0.9955(0.0945)	-1.008(0.1048)	-1.008(0.1048)	-1.0132(0.1041)
		β_{21}		-1.2866(0.1026)	-1.303(0.1126)	-1.303(0.1126)	-1.3084(0.1095)
		$\beta_{21} - \beta_{20}$		-0.2911(0.1063)	-0.2954(0.1232)	-0.2954(0.1232)	-0.3063(0.1215)
Type I	1		0.03	0.16	0.04	0.03	0.055
	2		0.07	0.095	0.055	0.055	0.077
	Combined						0.070
Estimate(s.e.)	1	β_{20}		-1.0257(0.0939)	-1.0758(0.1183)	-1.0178(0.13647)	-1.0296(0.1342)
		β_{11}		-1.0107(0.09337)	-1.0611(0.1179)	-1.0040(0.13568)	-1.0127(0.1321)
		$\beta_{11} - \beta_{10}$		0.0150(0.1059)	0.1179(0.1429)	0.0138(0.1754)	0.0169(0.1719)
	2	β_{20}		-0.9983(0.0946)	-1.0102(0.1046)	-1.0102(0.1046)	-1.0152(0.1031)
		β_{21}		-0.9888(0.0946)	-0.999(0.1045)	-0.999(0.1045)	-1.0107(0.1031)
		$\beta_{21} - \beta_{20}$		0.0095(0.1056)	0.0103(0.1221)	0.0103(0.1221)	0.0020(0.1220)

Appendix C

Supplementary Materials of Chapter 7

In this appendix, we present the code used, derivation of the formulas and additional tables for the simulation studies discussed in Chapter 7.

C.1 SAS code for three level hierarchical structure of the comet data

```
proc nlmixed data=aa  maxiter=5000 qpoints=30;
title 'Weibull normal normal for Tail Length';
parms int0=-28.7  int125=-40  int25=-41 int5=-42 int200=-38
rho=0 sig1=0 sig2=0;
if (dose=0) then eta=int0+b+b1*(slide=1)+b2*(slide=2)
+b3*(slide=3);
else if (dose=1.25) then eta=int125+b+b1*(slide=1)+b2*(slide=2)
+b3*(slide=3);
else if (dose=2.5) then eta=int25+b+b1*(slide=1)+b2*(slide=2)
+b3*(slide=3);
else if (dose=5) then eta=int5+b+b1*(slide=1)+b2*(slide=2)
+b3*(slide=3);
else if (dose=200) then eta=int200+b+b1*(slide=1)+b2*(slide=2)
+b3*(slide=3);
```

```
k= exp(eta);
erho=exp(rho);
loglik=rho+log(response)*(erho-1)+eta-((response**erho)*k);
esig1= exp(sig1);
esig2= exp(sig2);
model response ~ general (loglik);
random b b1 b2 b3 ~ normal([0,0,0,0],
[esig1,0,esig2,0,0,esig2,0,0,0,esig2]) subject=rat;
run;
```

Appendix D

Supplementary Materials of Chapter 10

In this appendix, additional tables for the first batch of simulation discussed in Chapter 10 are presented.

D.1 Result For the First Batch of Simulations

Table D.1: *Site as fixed effect; likelihood.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	959	24	8999	215	9958	239	0.024
2	0	MAR	Yes	0.25	2516	60	7526	175	10042	235	0.023
3	0	MAR	Yes	0.50	5590	157	4641	120	10231	277	0.027
4	0	MAR	No	-					9671	249	0.026
5	0	MAR	No	-					11380	300	0.026
6	0	MNAR	Yes	0.10	906	23	8975	226	9881	249	0.025
7	0	MNAR	Yes	0.25	2997	71	6986	158	9983	229	0.023
8	0	MNAR	Yes	0.50	5396	146	4659	105	10055	251	0.025
9	0	MNAR	No	-					9616	223	0.024
10	0	MNAR	No	-					11354	297	0.026
11	0	No	No	-					10158	213	0.021
12	0.2	MAR	Yes	0.10	1006	20	9040	269	10046	289	0.029
13	0.2	MAR	Yes	0.25	2448	69	7553	211	10001	280	0.028
14	0.2	MAR	Yes	0.50	5532	149	4510	139	10042	287	0.029
15	0.2	MAR	No	-					9616	234	0.024
16	0.2	MAR	No	-					11416	287	0.025
17	0.2	MNAR	Yes	0.10	932	21	8947	223	9879	244	0.025
18	0.2	MNAR	Yes	0.25	2946	65	6864	162	9810	227	0.023
19	0.2	MNAR	Yes	0.50	5435	148	4631	93	10066	241	0.024
20	0.2	MNAR	No	-					9574	235	0.025
21	0.2	MNAR	No	-					11430	288	0.026
22	0.2	No	No	-					10185	245	0.024

Table D.2: *Site as fixed effect; LOCF.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	923	923	8045	8045	8968	8968	1
2	0	MAR	Yes	0.25	2298	2298	6684	6684	8982	8982	1
3	0	MAR	Yes	0.50	4913	4913	4119	4119	9032	9032	1
4	0	MAR	No	-					9909	9909	1
5	0	MAR	No	-					8784	8784	1
6	0	MNAR	Yes	0.10	891	891	8058	8058	8949	8949	1
7	0	MNAR	Yes	0.25	2416	2416	6724	6724	9140	9140	1
8	0	MNAR	Yes	0.50	4894	4894	4096	4096	8990	8990	1
9	0	MNAR	No	-					9776	9776	1
10	0	MNAR	No	-					8731	8731	1
11	0	No	No	-					10158	213	0.021
12	0.2	MAR	Yes	0.10	939	939	7961	7961	8900	8900	1
13	0.2	MAR	Yes	0.25	2326	2326	6708	6708	9034	9034	1
14	0.2	MAR	Yes	0.50	4983	4983	4010	4010	8993	8993	1
15	0.2	MAR	No	-					9916	9916	1
16	0.2	MAR	No	-					8793	8793	1
17	0.2	MNAR	Yes	0.10	874	874	8222	8222	9096	9096	1
18	0.2	MNAR	Yes	0.25	2770	2770	6350	6350	9120	9120	1
19	0.2	MNAR	Yes	0.50	4910	4910	4119	4119	9029	9029	1
20	0.2	MNAR	No	-					10040	10040	1
21	0.2	MNAR	No	-					8951	8951	1
22	0.2	No	No	-					245	10185	0.024

Table D.3: *Site as random effect.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	997	22	9003	229	10000	251	0.025
2	0	MAR	Yes	0.25	2551	56	7449	191	10000	247	0.025
3	0	MAR	Yes	0.50	5527	150	4473	122	10000	272	0.027
4	0	MAR	No	-					10000	220	0.022
5	0	MAR	No	-					10000	288	0.029
6	0	MNAR	Yes	0.10	954	21	9046	220	10000	241	0.024
7	0	MNAR	Yes	0.25	2993	75	7007	171	10000	246	0.025
8	0	MNAR	Yes	0.50	5402	146	4598	99	10000	245	0.025
9	0	MNAR	No	-					10000	222	0.022
10	0	MNAR	No	-					10000	269	0.027
11	0	No	No	-					10000	261	0.026
12	0.2	MAR	Yes	0.10	990	24	9010	256	10000	280	0.028
13	0.2	MAR	Yes	0.25	2550	65	7450	177	10000	242	0.024
14	0.2	MAR	Yes	0.50	5537	143	4463	105	10000	248	0.025
15	0.2	MAR	No	-					10000	231	0.023
16	0.2	MAR	No	-					10000	221	0.022
17	0.2	MNAR	Yes	0.10	929	32	9071	249	10000	281	0.028
18	0.2	MNAR	Yes	0.25	2988	57	7012	158	10000	215	0.022
19	0.2	MNAR	Yes	0.50	5410	132	4590	89	10000	221	0.022
20	0.2	MNAR	No	-					10000	219	0.022
21	0.2	MNAR	No	-					10000	247	0.025
22	0.2	No	No	-					10000	274	0.027

Table D.4: *Site as random effect.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	997	22	9003	229	10000	251	0.025
2	0	MAR	Yes	0.25	2551	56	7449	191	10000	247	0.025
3	0	MAR	Yes	0.50	5527	150	4473	122	10000	272	0.027
4	0	MAR	No	-					10000	220	0.022
5	0	MAR	No	-					10000	288	0.029
6	0	MNAR	Yes	0.10	954	21	9046	220	10000	241	0.024
7	0	MNAR	Yes	0.25	2993	75	7007	171	10000	246	0.025
8	0	MNAR	Yes	0.50	5402	146	4598	99	10000	245	0.025
9	0	MNAR	No	-					10000	222	0.022
10	0	MNAR	No	-					10000	269	0.027
11	0	No	No	-					10000	261	0.026
12	0.2	MAR	Yes	0.10	990	24	9010	256	10000	280	0.028
13	0.2	MAR	Yes	0.25	2550	65	7450	177	10000	242	0.024
14	0.2	MAR	Yes	0.50	5537	143	4463	105	10000	248	0.025
15	0.2	MAR	No	-					10000	231	0.023
16	0.2	MAR	No	-					10000	221	0.022
17	0.2	MNAR	Yes	0.10	929	32	9071	249	10000	281	0.028
18	0.2	MNAR	Yes	0.25	2988	57	7012	158	10000	215	0.022
19	0.2	MNAR	Yes	0.50	5410	132	4590	89	10000	221	0.022
20	0.2	MNAR	No	-					10000	219	0.022
21	0.2	MNAR	No	-					10000	247	0.025
22	0.2	No	No	-					10000	274	0.027

Table D.5: *Site excluded from the final analysis.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	987	2	9013	6	10000	8	0.0008
2	0	MAR	Yes	0.25	2580	1	7420	4	10000	5	0.0005
3	0	MAR	Yes	0.50	5553	2	4447	5	10000	7	0.0007
4	0	MAR	No	-					10000	5	0.0005
5	0	MAR	No	-					10000	18	0.0018
6	0	MNAR	Yes	0.10	956	0	9044	6	10000	6	0.0006
7	0	MNAR	Yes	0.25	3033	0	6967	4	10000	4	0.0004
8	0	MNAR	Yes	0.50	5446	1	4554	5	10000	6	0.0006
9	0	MNAR	No	-					10000	6	0.0006
10	0	MNAR	No	-					10000	19	0.0019
11	0	No	No	-					10000	2	0.0002
12	0.2	MAR	Yes	0.10	984	2	9016	5	10000	7	0.0007
13	0.2	MAR	Yes	0.25	2545	3	7455	7	10000	10	0.0010
14	0.2	MAR	Yes	0.50	5532	3	4468	5	10000	8	0.0008
15	0.2	MAR	No	-					10000	2	0.0002
16	0.2	MAR	No	-					10000	20	0.0020
17	0.2	MNAR	Yes	0.10	945	2	9055	5	10000	7	0.0007
18	0.2	MNAR	Yes	0.25	2985	2	7015	5	10000	7	0.0007
19	0.2	MNAR	Yes	0.50	5434	3	4566	3	10000	6	0.0006
20	0.2	MNAR	No	-					10000	1	0.0001
21	0.2	MNAR	No	-					10000	15	0.0015
22	0.2	No	No	-					10000	0	0

Table D.6: *Site excluded in generating the data and in the final analysis.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	993	23	9007	236	10000	259	0.026
2	0	MAR	Yes	0.25	2563	64	7437	209	10000	273	0.027
3	0	MAR	Yes	0.50	5520	144	4480	121	10000	265	0.027
4	0	MAR	No	-					10000	246	0.025
5	0	MAR	No	-					10000	258	0.026
6	0	MNAR	Yes	0.10	952	22	9048	237	10000	259	0.026
7	0	MNAR	Yes	0.25	2991	69	7009	192	10000	261	0.026
8	0	MNAR	Yes	0.50	5397	138	4603	107	10000	245	0.025
9	0	MNAR	No	-					10000	237	0.024
10	0	MNAR	No	-					10000	254	0.025
11	0	No	No	-					10000	244	0.024
12	0.2	MAR	Yes	0.10	990	35	9010	228	10000	263	0.026
13	0.2	MAR	Yes	0.25	2556	82	7444	192	10000	274	0.027
14	0.2	MAR	Yes	0.50	5476	158	4524	119	10000	277	0.028
15	0.2	MAR	No	-					10000	239	0.024
16	0.2	MAR	No	-					10000	255	0.026
17	0.2	MNAR	Yes	0.10	948	29	9052	225	10000	254	0.025
18	0.2	MNAR	Yes	0.25	2981	96	7019	177	10000	273	0.027
19	0.2	MNAR	Yes	0.50	5389	142	4611	119	10000	261	0.026
20	0.2	MNAR	No	-					10000	232	0.023
21	0.2	MNAR	No	-					10000	253	0.025
22	0.2	No	No	-					10000	242	0.024

Samenvatting

Dit proefschrift is een studie naar specifiek modellen en design aspecten voor toxicologische studies en klinische proeven. De tekst is opgebouwd uit twee delen.

Het eerste deel van het proefschrift wordt gemotiveerd door een toxicologische studie voor het opmeten van DNA schade: het zogenaamde comet assay. De studie werd ontwikkeld door Ostling en Johansson in 1984 en werd later aangepast door Sing *et al* in 1988. De populariteit van deze methode is sindsdien steeds toegenomen voor de evaluatie van DNA schade en toetsen voor genotocixiteit, en wordt nu gebruikt als een standaard methode voor de beoordeling van de veiligheid van potentiële nieuwe geneesmiddelen. DNA migratie wordt in de comet assay gevisualiseerd (typisch komeet-achtige structuur), welk toelaat de DNA schade te kwantificeren op het niveau van de cel. Drie maten worden hier meestal voor gebruikt: staart migratie (of staartlengte van de komeet), percentage staart intensiteit en staart-moment. Verschillende eigenschappen maken de analyse van deze data moeilijk: (1) Het ontwerp van de studie is hiïarchisch in de zin dat dieren genest zijn in dosisgroepen, een aantal 'slides' per dier gebruikt worden en verschillende cellen gemeten worden per 'slide'. (2) De opmetingen van de studie zijn niet normaal verdeeld maar eerder asymmetrisch, scheef, bi- of multimediaal, een mengvorm van verdelingen, enz. Traditionele analyses negeren geheel of gedeeltelijk de eigenschappen van de gegevens, en zijn gebaseerd op een samenvatting van de metingen per cluster. In dit deel worden statistische modellen onderzocht welke rekening houden met de hiërarchische structuur en overdispersie in de data, zoals voor de comet essay gegevens. Zowel uni- als multivariabele modellen worden beschouwd. Ter wille van de computationele complexiteit van de voorgestelde modellen, wordt een alternatieve schattingsmethode onderzocht.

De standaard aanpak voor het modelleren van niet-normaal verdeelde gegevens is het gebruik van veralgemeende lineaire modellen. De meest bekende voorbeelden zijn lineaire regressie, logistische regressie en Poisson regressie. Een belangrijke uit-

breiding van deze modellen is het veralgemeende lineaire gemengde model, door het toevoegen van een normaal-verdeeld random effect om de hiërarchische structuur van de data in rekening te brengen. Echter, een veel voorkomend probleem met niet-normale data is overdispersie in de zin dat de variabiliteit in de data niet voldoet aan de gemiddelde-variantie relatie afkomstig van de verdeling. Dit komt zowel voor in het geval van univariate als multi-level data. In het geval van een veralgemeend lineair model maakt men vaak gebruik van een geconjugeerd random effect om rekening te houden met overdispersie, zoals in het negatief binomiaal model of het beta-binomiaal model. In een recente publicatie stelden Molenberghs, Verbeke en Demétrio (2007) een gelijkaardige methode voor om rekening te houden met overdispersie in een multivariate setting, door het gebruik van twee random effecten, een normaal verdeeld random effect om rekening te houden met de hiërarchie, en een geconjugeerd random effect om rekening te houden met overdispersie in de data. Dit wordt vaak aangeduid als het *gecombineerde model*.

In Hoofdstuk 4 werd een flexibele modelleer-methode voorgesteld voor hiërarchisch geclusterde niet-normale data met overdispersie, in het kader van het comet essay. Het model maakt gebruik van de volledige hiërarchische structuur, de meest geschikte verdeling, en overdispersie, via het gebruik van het gecombineerde model. In deze methode maken we gebruik van een geconjugeerd gamma random effect om rekening te houden met overdispersie, en een gamma en normaal random effect om rekening te houden met de hiërarchie. De meer conventionele modellen welke enkel rekening houden met ofwel overdispersie ofwel een enkel hiërarchisch random effect worden beschouwd als sub-modellen. Het biedt een ruime keuze aan modellen bij model selectie. Het gebruik van meer uitgebreide modellen met overdispersie en de hiërarchische structuur was een verbetering in deze studie.

Tot op dit punt werden de resultaten univariaat gemodelleerd. De comet assay data vertonen echter een multivariate structuur. Ook in vele andere klinische studies en toxicologische experimenten komen multivariate longitudinale of geclusterde data voor. In toxicologie is er meestal geen enkele standaard eindpunt om de toxiciteit en de werkzaamheid van de verbinding te evalueren, maar worden co-primaire eindpunten gebruikt om de toxiciteit te beoordelen. Het gezamenlijk modelleren van deze uitkomsten is dus aantrekkelijk voor algemene besluitvorming en om associaties op te meten. Een verdere uitbreiding naar een multivariate setting met hiërarchisch geclusterd niet-normale uitkomsten met overdispersie wordt voorgesteld in Hoofdstuk 5. Beide resultaten worden gezamenlijk geanalyseerd door te veronderstellen dat de normale verdeling van de random effecten gecorreleerd zijn. De associatie structuur

tussen de uitkomsten werd analytisch afgeleid.

In Hoofdstuk 6 worden alternatieve verdeling verondersteld voor de overdispersie: (1) Een discrete verdeling welke uiteindelijk leidt tot een eindige mengeling van modellen om rekening te houden met overdispersie. (2) 'Zero-inflated' modelled worden onderzocht om rekening te houden met de toename aan nullen. (3) Een mengeling van geconjugeerde verdeling, welke toelaten afwijkingen van een enkele geconjugeerde verdeling in rekening te brengen, terwijl de eigenschap van geconjugeerd zijn toch gebruikt kan worden om de berekeningen te vereenvoudigen. Dit verhoogt de flexibiliteit van het gecombineerde model.

De meeste van de statistische modellen beschouwd in deze thesis zijn gebaseerd op het gecombineerde model van Molenberghs *et al.* (2010). Het voornaamste probleem bij dit soort modellen is de computationeel complexe schatting ter wille van de multivariate integralen, net zoals bij veralgemeende lineair gemengde modellen. In Hoofdstuk 7 wordt het gebruik van Gaussische variationele benadering onderzocht als een alternatieve schattingsmethode. In deze methode wordt de integrand benaderd door het invoeren van een set van variationele dichtheden, zodat evaluatie van de integrand eenvoudig is. De methode is erg nuttig indien er computationele problemen optreden met standaard schattingsmethoden, of wanneer rekentijd te groot is.

Het tweede deel van deze thesis heeft betrekking op onvolledige gegevens in klinische studies. In klinische studies, hoewel de primaire focus van veel onderzoek op een specifieke tijd van de meting is, meestal het laatste tijdstip, wordt de uitkomst opgevolgd in tijd, en onvolledige gegevens en uitval in het bijzonder komen vaak voor. Deze ontbrekende gegevens hebben in het algemeen een invloed en beïnvloeden de gevolgtrekking. Dit kan leiden tot vertekende resultaten en kan een ernstig verlies van onderscheidend vermogen opleveren als het aandeel van de onvolledigheid hoog is. Sommige experimenten laten een data-gedreven aanpassing toe wanneer de uitval groot is. Bijvoorbeeld, als de uitval groot is in één van de behandelingsgroepen, kan deze behandelingsgroep uit de studie gelaten worden en wordt een andere behandelingsgroep met minder ontbrekende data in de plaats daarvan gebruikt. Dus, men laat toe om de vergelijking tussen behandelingen te wijzigen aan het einde van de studie. De betrouwbaarheid van de inferentie kan in vraag gesteld worden, aangezien dit een impact heeft op de type I fout. In Hoofdstuk 10 wordt de impact van dergelijke data-gestuurde aanpassingen onderzocht. In het bijzonder wordt de type I fout geassocieerd met het wisselen van dosis groepen onderzocht als de primaire uitkomst longitudinaal is. Een verwachting is dat de type I fout zal toenemen door het wis-

selen van dosis groep, en dat significantieniveau aangepast moet worden. Langs de andere kant is het mogelijks niet van toepassing in andere studies, zoals de beschouwde studie op de ziekte van Alzheimer. De Type I fout kan kleiner zijn voor dergelijke studies, dus geen toename van de algemene Type I fout, indien de primaire analyse aangepast werd door het wisselen van dosis groepen gebaseerd op hoge uitval in de hoge dosis groep. In deze studie werd een simulatie uitgevoerd om de type I fout te onderzoeken, naar analogie met de studie naar de ziekte van Alzheimer. De Type I fout werd onderzocht in verschillende scenario's, in functie van verschillende correlaties tussen effectiviteit en tolerantie, verschillende mechanisme voor ontbrekende data, en verschillende kansen op te wisselen. Een reeks van parameterwaarden werd gebruikt om de sensitiviteit van de analyse te bestuderen. We besluiten dat, hoewel het wisselen van dosissen op een data-gestuurde manier in de finale analyse, waar het criteria voor wisselen gerelateerd is met primaire uitkomst, de type I fout in het algemeen zal doen toenemen, de toename van de type I fout onder controle is voor de meeste scenario's van de Alzheimer studie met longitudinale uitkomsten waar patiënten verwacht worden te verslechteren met de tijd. Een uitzondering werd gezien voor het specifieke geval waar uitval sterk geassocieerd is met de effectiviteit. De criteria gebruikt voor wijziging moet bijgevolg zorgvuldig onderzocht worden met betrekking tot de associatie met significantie van het behandelingseffect.