KU LEUVEN

FACULTY OF SCIENCE

# Exploration of Extended Models for Overdispersed, Repeated Time-to-Event Data

**Martial LUYTS**

Supervisor: Prof. dr. G. Molenberghs
Affiliated at *the faculty of Medicine of the Catholic University of Leuven* & *the faculty of Science of the Hasselt University*

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Statistics

Academic year 2014-2015

# Acknowledgements

For completing the masters in Statistics at KULeuven, students have to write an own research topic, related to statistics. The main purpose of this topic is to actively contribute to a scientific research of own interest. Before and during my two year masters education in Leuven, I gained an increased interest in longitudinal and survival data analysis. Searching for a combination between both analytical approaches, and collaborating ideas with Prof. dr. Molenberghs, resulted in a topic that was related to previously done research and current research developments. Additional, lot of mathematical derivations were involved, which support my interest in mathematical statistics. In this part, I would like to acknowledge very important persons who have contributed through diverse ways to the success of this thesis.

First of all, many thanks is given to my promotor, Prof. dr. Geert Molenberghs. During my academic career at KULeuven and UHasselt, he has been an amazing professor who inspired me in proceeding research in one of his own research topics. The first time that I met Geert, other professors told me that he is one of the best, a world phenomena, in longitudinal data analysis. After three beautiful years of following classes of him, and collaborating ideas for my bachelor and master thesis, I am able to say: "I fully agree on this statement!". He gave me all the necessary feedback and insights during my bachelor- and masterthesis, and was always there when questions and problems arises. Many thanks for the wonderful years!

Besides Prof. dr. Molenberghs, I also would like to thank dr. Edmund Njagi, for providing relevant mathematical materials and a nice discussion about joint modeling, dr. Samuel Iddi and Trias Wahyuni Rakhmawati, for providing relevant SAS code, and jury members Prof. dr. Geert Verbeke and Prof. dr. Emmanuel Lesaffre, for providing and still making amazing research about this thesis topic. Thanks is also given to all other statistical and mathematical professors from KULeuven and UHasselt.

Furthermore, special thanks is given on two of my best friends, i.e., Kris Peremans and Caro Lievens. During my university career, they helped me a lot in mathematics and statistics, and have been best friends ever since. By this way, I would like to say: "Thanks guys, you are amazing!".

This thesis will not have been possible without the support of my family. To my father and mother, grateful acknowledge is provided for the continuous support, assistance and encouragement throughout my education. They don't only gave me the opportunity to study at the university, they also have been a reliable and supporting source when difficulties arises.

At last, I greatly acknowledge KULeuven and UHasselt for the interesting courses!

# Abstract

Non-Gaussian outcomes are often modeled using members of the so-called exponential family. Notorious members are the Bernoulli model for binary data, leading to the classical logistic regression, and the Exponential model for time-to-event data. However, limitations arises when complex data structures are present, indicating a need for extending this family. Two main reasons for this extension are (1) the occurrence of overdispersion, meaning that the variability in the data is not adequately described by the models, which often exhibit a prescribed mean-variance link, and (2) the accommodation of hierarchical structure in the data, stemming from clustering in the data which, in turn, may result from repeatedly measuring the outcome, for various members of the same family, etc. To accommodate both issues simultaneously, the framework of Molenberghs et al (2010) is proposed, where both complexities are accommodated through two separate sets of random effects, i.e., the so-called conjugate random effects at the level of the mean for the first aspect and (2) normal random effects embedded within the lineair predictor for the second aspect. Apart from model formulation, generic approximations for marginal model elements, e.g., the marginal mean, variance and covariance, and the strong conjugacy principle are explored.

Over the years, massive attention has grown for creating new time-to-event modeling structures. Since an extra complexity arises in this setting, i.e., censoring, Molenberghs et al (2014) included gamma and normal random effects in a Weibull model, to account for overdispersion and between-subject effects, respectively, and additionally accounted for censoring within two estimation methods, i.e., maximum likelihood with partial marginalization and pseudo-likelihood. In this thesis, full attention is given on their proposed methodology. Next to these fixed effects estimation, empirical bayes estimation are used to explore estimation for the random effects.

Due to the flexible modeling structure, more complex clustering formats can easily be taken into account with the use of the alternating imputation posterior algorithm (Efendi and Molenberghs, 2013). Even though subject-specific interpretations are retrieved from the fixed effects within the framework, population-averaged interpretations can still be achieved by the use of a connector function. Furthermore, joint modeling structures can easily be made within the framework, even in such a way that informative censoring, i.e, an aspect that was quite difficult to take into account in the past, can be captured in the modeling framework. Going even deeper in the setting of joint modeling, a conceptual correspondence exists between the missing data setting, and joint modeling of longitudinal and time-to-event outcomes (Njagi et al, 2013c), which makes the methodology quite interesting to many researchers.. A characterization of missing at random is provided within the missing data setting.

Equivalent to other classical methodologies, diagnostic tools were (and still are) development for the framework. First, the gradient function of Verbeke and Molenberghs (2013) is discussed and used as simple graphical exploratory diagnostic tool to assess whether the assumed random-effects distribution produces an adequate fit to the data, in terms of marginal likelihood. While no additional computations, other than the computations needed to fit the model, are required, its applicable to a wide range of models within the framework, as long as the distribution for the outcomes conditional on the random effects is correctly specified. Moreover, the function gives an indication on how a parametric model can be improved in case of misspecification. Based upon the gradient function, Efendi, Drikvandi, Verbeke and Molenberghs (2014) developed a simple diagnostic test for the random-effects distribution in mixed models, where the function serves as basis for the construction of the proposed formal test. Secondly, the detection of influential observations is explored by the local influence paradigm of Rakhmawati, Molenberghs, Verbeke and Faes (2014). Main advantage of this approach is the ease of interpretable and computationally convenient expressions, not only highlighting influential subjects, but also which aspect of their profile leads to undue influence on the model's fit (Verbeke and Lesaffre, 1998).

To end the theoretical discussion of this thesis, emphasis is placed on two principal ways in which the proposed model extend beyond the data available, i.e., (1) the data may be coarsened, i.e., what is actually observed is less detailed than what is planned, and (2) the data may be augmented, i.e., the observed data are hypothetically but conveniently supplemented with structures. A typically cause for the former one can be censoring, while the latter structure can be random effects, implying that its reasonable to discuss this for the proposed framework for time-to-event data. Both aspects together are referred as enriched data. The fitting of such methodologies combines evidence arising from empirical data with non-verifiable model components, i.e., that are purely assumption driven. Therefore, discretion of the potential dangers and pitfalls that follow from this should be present in the analysis, indicating the importance of sensitivity analysis, i.e., a methodology that studies how assumptions about unobservables, given the observables, influence the inferences drawn. Moreover, to any given model, an entire class of models can be assigned, with all members producing the same fit to the observed data but arbitrary regarding the unobservable parts of the enriched data.

The methodology is applied to survival data in children with asthma. A full analysis is provided, in a way that all theoretical discussed aspects within the methodology are conveniently applied and explored in detail. While some have already been analyzed in previously published papers, others were totally new in research. Code and derivations are found in the appendix, while a general conclusion is provided at the end.

# Samenvatting

In statistiek komt het vaak voor dat uitkomsten een niet-normaal verdeeld patroon vertonen. Het analyseren van deze types wordt traditioneel gedaan met verdelingen uit de zo-genaamde exponentiële familie, waarbij de Poisson voor aantallen en de Bernoulli verdeling voor binaire data de meest gekende zijn. Door de aanwezigheid van complexere structuren kan het voorkomen dat een slechte schatting wordt gevormd met deze modellen. De nood aan uitbreidingen is daarom voorhanden. Twee redenen die een belangrijke rol spelen bij deze conclusie zijn (1) de aanwezig van overdispersie, betekende dat de variabiliteit in de data niet adequaat wordt beschreven door de modellen, welke vaak een bepaalde gemiddelde-variantie link bevat, en (2) de mogelijke hiërarchische structuur in de data, afkomstig door clusteringen in de data, welke op hun beurt, resulteert uit herhaalde metingen van de uitkomst.

In het verleden is veel onderzoek verricht naar het vinden van gepaste, uitgebreidere modelleringstechnieken die beide aspecten in rekening brengen. Terwijl deze twee aspecten vaak afzonderlijk werden behandeld (Hinde en Demétrio, 1998ab; Verbeke et al, 2000; Molenberghs et al, 2005), ontwikkelde Molenberghs et al (2010) een elegant raamwerk dat de mogelijkheid biedt om beide aspecten gelijktijdig in rekening te brengen d.m.v. twee afzonderlijke verzamelingen van random effecten, en zelfs apart te behandelen. Deze twee random-effecten, ook wel conjugate en normale random-effecten genoemd, brengen overdispersie en hiërarchische structuren in rekening door een directe imputatie op het gemiddelde-level en binnen de lineaire predictor, respectievelijk. Naast de formuleringen van het raamwerk, werden ook generieke benaderingen van de marginale elementen, zoals het marginaal gemiddelde en variantie, behandeld, alsook het principe van sterke conjugatie.

In deze thesis wordt specifieke focus gelegd op overlevingstijden. Aangezien de overlevingstijden, vaak gemodelleerd door de exponentiële verdeling (element uit de exponentiële familie), een extra complexiteit met zich meebrengen, namelijk censurering, is het aangeraden om extra adaptaties te verrichten binnen het raamwerk van Molenberghs et al (2010), die het mogelijk maken om censurering in rekening te brengen. Met deze gedachtengang in plaats creeërde Molenberghs et al (2014) een raamwerk dat zowel overdispersie, hiërarchische structuren en censurering in rekening bracht door het raamwerk van Molenberghs et al (2010) te adopteren, en censurering apart behandeld in twee schattingsmethoden: (1) maximum likelihood met partiële marginalisatie en (2) pseudo-likelihood. Dit framework, waarbij de Weibull-Gamma-Normal model vooreerst vermeld werd, wordt gebruikt als basis beginsel in de discussie rond overdisperste, herhaalde overlevingstijden. Naast het behandelen van de schattingstechnieken rond vaste effecten, is de "empirical bayes" schatting voorgesteld voor het schatten van de persoons-gebonden random effecten.

Door de flexibele structuur van het raamwerk zijn vele uitbreidingen mogelijk voor onderzoek. Bijvoorbeeld, complexere hiërarchische structuren kunnen gemakkelijk gemodelleerd worden, door toevoeging van extra random effecten aan het model. Terwijl schattingen steeds complexer en moeilijker worden voor stijgende random effecten, introduceerde Efendi en Molenberghs (2013) het gebruik van het "alternating imputation posterior" algoritme binnen deze setting. Verder, zowel persoons-gebonden interpretaties als populatie-gemiddeld interpretaties kunnen verkregen worden in het model, mits implementatie van een bepaalde connector functie. Terwijl men conventioneel persoons-gebonden interpretaties binnen het raamwerk krijgt, presenteerde Efendi, Molenberghs en Iddi (2014) een manier om populatie-gemiddelde interpretaties te bekomen binnen het framework, d.m.v. een connector functie. Het verhaal stopt hier zelfs niet. Naast univariate analyses, waarbij slechts 1 uitkomst wordt gemodelleerd, kunnen gezamenlijke analyses plaatsvinden, waarbij verschillende types van uitkomsten tegelijkertijd worden gemodelleerd. In een cardiologische studie, bijvoorbeeld, waarin onderzoekers gebruik maken van telemonitoring (een techniek waarmee patiënten vanop afstand worden gevolgd) om bloeddruk te meten op dagelijkse basis, worden vaak gecombineerd met additionele gegevens zoals hartslag en gewicht, naast tijd tot heropname. Gezamenlijke analyses kunnen gemakkelijk toegepast worden in het framework, zelfs op een manier dat informatieve censurering in rekening kan worden gebracht. Verder, Njagi et al (2013c) toonde aan dat een conceptueel verband bestaat tussen de missing data methodologie en het gezamenlijk modelleren van longitudinale uitkomsten en overlevingstijden, welke de methodologie interessant maakt in verschillende statistische settings. Om de discussie hierover te beëindigen, wordt een karakteristiek van "missing at random" gegeven binnen de missing data setting.

Traditioneel, na het formuleren en beschrijven van een modelleringstechniek, wordt er aandacht geschonken aan diagnostische testen. In context van het framework van Molenberghs et al (2014), worden zowel de gradient functie van Verbeke en Molenberghs (2013) als het lokale invloed paradigma van Rakhmawati, Molenberghs, Verbeke en Faes (2014) behandeld. Terwijl de eerste wordt gebruikt als grafisch verkennend diagnostisch hulpmiddel om te beoordelen of de veronderstelde random-effecten verdeling een voldoende schatting produceert in termen van marginale kans, wordt detectie van invloedrijke observaties onderzocht door de tweede. Voor de gradient functie zijn geen extra berekeningen nodig, anders dan de berekeningen om het model te schatten, en is toepasbaar op tal van modellen binnen het voorgesteld framework, zolang de verdeling op de uitkomsten conditioneel op de random effecten correct gespecificeerd is. Verder geeft deze een indicatie weer hoe het parametrisch model verbeterd kan worden in geval van misspecificatie. Gebaseerd op deze functie, is een eenvoudige diagnostische test voor de random-effecten verdeling voor gemengde modellen ontwikkeld (Efendi, Drikvandi, Verbeke en Molenberghs, 2014), waarbij de functie dient als basis voor de constructie van de voorgestelde formele toetsstatistiek. Voor het lokale invloed paradigma worden invloedrijke observaties geïdentificeerd door interpreteerbare en computationeel logische uitdrukkingen. Verder geeft de procedure weer welke aspecten in hun profiel leid tot invloedrijke resultaten in de modelschatting (Verbeke and Lesaffre, 1998).

Om de theoretische discussie rond het framework te beëindigen, wordt er specifieke nadruk gelegd op twee belangrijke manieren die samen zogenaamde verrijkte gegevens uitmaken. Coarsening is één van de twee klassen, in de zin dat de gegevens op een minder fijn niveau

worden opgemeten dan men in principe zou willen. Voorbeelden hiervoor zijn censurering en onvolledige gegevens. De andere, ook wel augmentatie genoemd, is waar aan gegevens structuren worden toegevoegd die niet worden geobserveerd doch het modelleren faciliteren. Voorbeelden zijn random effecten en latente veranderlijken. In alle gevallen is een deel van het model aangestuurd enkel door veronderstellingen en niet door gegevens. De gevaren daarvan dienen ten volle onderkend (Verbeke en Molenberghs, 2010; Molenberghs et al, 2012). Het is daarom goed om te concluderen dat er naast goodness-of-fit ook nood aan sensitiviteitsanalyse is.

Dit werk sluit af met een uitgebreide analyse van de astma dataset (Duchateau en Janssen, 2008), waarbij alle theoretische aspecten in detail worden behandeld, en een algemene conclusie rond het framework. Code en berekeningen worden ondergebracht in de appendix.

# List of Abbreviations

| | |
|---|---|
| **AIP** | Alternating Imputation Posterior |
| **CHF** | Chronic Heart Failure |
| **CM** | Combined Model |
| **COMMM** | Combined Overdispersed and Marginalized Multilevel Model |
| **EB** | Empirical Bayes |
| **EM** | Expectation-Maximization |
| **ESPM** | Extended Shared Parameter Model |
| **GEE** | Generalized Estimation Equations |
| **GLM** | Generalized Linear Model |
| **GLMM** | Generalized Linear Mixed Model |
| **GSPM** | Generalized Shared-Parameter Model |
| **ICC** | Intraclass Correlation |
| **IMT** | Information Matrix Test |
| **JMMM** | Joint Marginalized Multilevel Model |
| **LBN** | Logit-Beta-Normal |
| **LMM** | Linear Mixed Model |
| **LSE** | Least Squares Estimation |
| **MAR** | Missing At Random |
| **MCAR** | Missing Completely At Random |
| **MCMC** | Markov Chain Monte Carlo |
| **MIMT** | Modified Information Matrix Test |

| | |
|---|---|
| **MLE** | Maximum Likelihood Estimation |
| **MMM** | Marginalized Multilevel Model |
| **MNAR** | Missing Not At Random |
| **MQL** | Marginal Quasi-Likelihood |
| **PBN** | Probit-Beta-Normal |
| **PMM** | Pattern-Mixture Model |
| **POMM** | Proportional Odds Mixed Model |
| **PQL** | Penalized Quasi-Likelihood |
| **RMSD** | Root Mean Squared Deviation |
| **SEM** | Selection Model |
| **SET** | Sandwich Estimation Test |
| **SNP** | Semi-nonparametric |
| **SPM** | Shared-Parameter Model |
| **SSE** | Sum of Squared Errors |
| **WGMNM** | Weibull-Gamma-Multivariate-Normal Model |
| **WGN** | Weibull-Gamma-Normal |
| **WGNN** | Weibull-Gamma-Normal-Normal |
| **WN** | Weibull-Normal |

# List of Tables

# List of Figures

# Contents

# Chapter 1

# General Introduction

## 1.1 Introduction

Life science studies often encompasses extended structures such as longitudinal collected data, where subjects/patients are repeatedly measured over time, and hierarchical structures, originating from hierarchical designs, e.g., multi-centre trials, surveys with a geographical-hierarchical framework such as the Belgian Health interview, which have extensively been studied in the past (Van Oyen et al, 1997; Van Oyen and Tafforeau, 1994). Keeping these structures aside in statistical modeling frameworks are often insufficiently in drawing valid conclusions, indicating the need of a modeling framework(s) that (perfectly) reflects the design of the study and results in much more reliable conclusions, for survey research, surrogate marker and surrogate endpoint evaluation, clinical trials, etc.

Therefore, its suggestive to search for a flexible modeling framework that supersedes a variety of currently available frameworks for complex data structures such as hierarchies. Simple and user-friendly implementation strategies are advisable in standard software packages, where model assessment and model diagnostics are available and simple to implement. Principle treatments of incomplete data, especially when targeting human study subjects, need to be available in the exploration, since life science studies are often connected with lots of missing data. The framework should be able to model different outcome types such as count, binary, continuous and time-to-event data, and, in case of non-Gaussian outcomes, accommodate typical design features such as overdispersion, zero-inflation, etc.

In the past, many modeling approaches have been proposed, which often can be placed within the generalized linear modeling (GLM) framework (Nelder and Wedderburn 1972, McCullagh and Nelder 1989, Agresti 2002), i.e., an unifying framework based on the so-called exponential family distributions. While this framework is mainly (not for the normally distributed outcomes, where the mean and variance are entirely separated from each other) restricted with the so-called mean-variance relationship, i.e., the variance is expressed as a deterministic function of the mean, extended frameworks need to be developed that accommodates this restriction. For example, Hammami, Garcia and Nuel (2013) showed that field-collected count data such as the number of parasites and the number of leukocytes per high power field are inconsistent with the Poisson assumptions. Failure to take this inconsistency into account in parasite and leukocyte counts may entail

important misleading inferences when these data are related to other explanatory variables (malariometric or environmental). Here, the inconsistency is referred as overdispersion, i.e., the observed variation is greater than predicted by the model. However, it should be noted that underdispersion, i.e., the observed variation is lower than predicted by the mode, can occur as well.

In case of count and binomial data, for example, overdispersion may be caused by increased in variation due to the omission of important unobserved covariates or the violation of independent assumption when observations are correlated or are collected from clusters. In case of purely binary data, hierarchies such as repeated measures or longitudinal data need to be present in the data structure in order to violate the mean-variance link. In practice, models like the negative-binomial model (Breslow, 1984; Lawless, 1987; Gardner et al, 1995; Engel, 1984; Manton et al, 1981) and zero-inflated Poisson (ZIP) model (Lambert, 1992; Hall, 2000) are often conducted for count data to overcome overdispersion.

Apart from the occurrence of over- or underdispersion, hierarchical structures imply the presence of association between measurements on the same unit as well. Therefore, alongside with the mean and variance function, the model should also include a proper reflection of the association structure within the data. Aspects like between and within subject-specific associations are crucial here. For non-Gaussian outcomes, the so-called generalized linear mixed model (GLMM, Engel and Keen 1994, Breslow and Clayton 1993, Wolfinger and OConnell 1993) has been suggested, and became a popular framework ever since in dealing with hierarchical data structures. In this methodology, random effects are introduced to capture the association structure and to some extent overdispersion. However, dealing with overdispersion and hierarchical structures separately may fall short when modeling the data (Molenberghs et al, 2007). Therefore, a combined modeling (CM) framework was set up that encompasses both aspects at the same time, by the inclusion of so-called conjugate random effects within the GLMM. Including all kinds of modeling settings in their framework, focus is laid on the time-to-event setting.

In time-to-event data, an extra complexity occurs within the data, i.e., censoring, either informative or non-informative. In the last decade, much research has been provided on non-informative censoring, meaning that participants who drop out of a study should do so due to reasons unrelated to the study. Informative censoring occurs when participants are lost to follow-up due to reasons related to the study, e.g. in a study comparing disease-free survival after two treatments for cancer, the control arm may be ineffective, leading to more recurrences and patients becoming too sick to follow-up. On the other hand, patients on the intervention arm can be completely cured by a treatment and may no longer feel the need to follow-up. While popular models in time-to-event data, e.g., the Cox proportional hazard model (Cox, 1972), almost invariable assume that the censoring is non-informative or ignorable, bias results often occur. In clinical trials, for example, it is often occurred that patients withdraw from the study. A possible reason is the better conditional state of the patient at that moment, such that no further medical attention is needed. In this case, the event that was proceeded before the censoring may have a significant effect, and may increase the expected remaining lifetime. Lagakos (1979) formulate an amount of examples where the assumption of non-informative censoring is doubtable.

To accommodate for non-informative censoring in the CM framework for time-to-event data, Molenberghs et al (2014) derived a strategy that keeps the CM formulation unharmed and, at the same time, accounts for censoring by simple imputing them within estimation approaches. However, unlike other frameworks, the CM framework also enables the ability to accommodate informative censoring in a certain way. Here, for example, informative censoring is taking into account in the CM framework for time-to-event data by the use of a specific joint multilevel model, where censoring is captured by specific correlated normal random effects.

Next to censoring, model interpretations differ. In the linear mixed model (LMM) for continuous outcomes, a marginal interpretation is present for the parameter estimates, even though a hierarchical has been employed. This is generally not the case for the GLMM for non-Gaussian outcomes, where the subject-specific random effect is used to capture between subject variability. Interpretation of fixed effect parameters are therefore conditional on subject-specific random effects. Since the CM framework holds the same reasoning, two different approaches can be explored, i.e., subject-specific/hierarchical and population-averaged/marginal interpretations. While subject-specific interpretations are often conducted in previously done research, Heagerty (1999) and Heagerty and Zeger (2000) proposed a so-called marginalized multilevel model (MMM), which combines the strength of the marginal and hierarchical models. Particularly, a direct marginal interpretations is present on the effect of covariates.

In the past decade, massive attention has grown in joint modeling, where several outcomes are modeled together in one particular model strategy. In life science applications, for example, this is not very uncommon. Researchers often collect several kinds of outcomes simultaneously in their studies, commonly of a mixed nature. For example, in toxicity studies, no single standard endpoint exists 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 these endpoints jointly not only appeals to draw overall inferences using all responses, it also captures the association among the endpoints, implying the importance in extending standard, univariate methodologies Therefore, various joint modeling approaches have been developed in the past. For longitudinal and time-to-event outcomes, a brief overview of methodologies can be found in Tsiatis and Davidian (2004).

After formulating a specific methodology, diagnostic tools are needed to explore the assessment of model fit. In case of the CM framework, different tools can be used. While one mainly focuses on testing the misspecification of the used random-effects proposition, influential observations can be explored as well. Model assessment diagnostics are crucial in deriving right conclusions, and need to be taken into account when modeling the data. For the CM framework, two very powerful tools have been developed in the last two years, i.e., (1) the gradient function of Efendi, Drikvandi, Verbeke and Molenberghs (2014), which serves as a graphical exploratory diagnostic tool to assess misspecification of the random effects distribution, and (2) the local influence paradigm of Rakhmawati, Molenberghs, Verbeke and Faes (2014), where influential observations can be detected in a simple manner.

This thesis attempts to address a number of these complexities of study designs and of the time-to-event data through the use of one specific member within the CM framework, i.e.,

the Weibull-Gamma-normal (WGN) model. Next to a theoretical discussion, the model is practically used to model the time to recurrence of an asthma attack in the asthma dataset of Duchateau and Janssen (2008).

## 1.2 Aim of this thesis

The general purpose of this thesis is to develop a modeling framework encompassing repeated, overdispersed time-to-event data, to offer some guidelines for future research. Additionally to the model formulation, emphasis is placed on the underlying theoretical study of the framework (e.g., to facilitate estimation and inferential methodology, even in higher dimensional hierarchical structures; the derivation of key features, such as the marginal parameters and correlation functions; the development of flexible joint hierarchical/marginalized models; formulating local influence diagnostics; developing goodness-of-fit tools for the random-effects distribution; and mentioning general attention for enriched-data problems and essential non-identifiability).

## 1.3 Outline

The thesis is organized as follows.

Motivating datasets are presented in Chapter 2. The aim of using this dataset is to demonstrate the practical use of the proposed modeling approaches. In Chapter 3, key ingredients of the modeling framework are considered in detail, where topics like the standard generalized linear model (in general case), extensions to overdispersion and mixed modeling, e.g., linear mixed model (LMM) and generalized linear mixed model (GLMM), are discussed in advance. These key ingredients are essential to understanding the basic idea behind the combined model (CM), introduced by Molenberghs et al (2007), which extends the classical GLMM by combining the normal random effects and the so-called conjugate random effects into a single framework to simultaneously address correlation in the hierarchical structures and overdispersion, respectively. Special attention of the (hierarchical) combined model in the field of time-to-event outcomes (with incorporation of censoring) is provided in Chapter 4, where a model formulation is given according to the so-called strong conjugacy principle. Emphasis is placed on a number of estimation strategies (e.g. maximum likelihood with partial marginalization (Molenberghs et al, 2007), pairwise likelihood (Molenberghs et al, 2014; Efendi et al, 2013) and Bayesian model fitting (Ghebretinsae et al, 2012) and coupled to the proposed combined model, even in combined models with higher order hierarchical structures (by using the so-called Alternating Imputation Posterior (AIP) algorithm (Efendi et al, 2013)).

In Chapter 5, we adapt the combined model from Chapter 4 to provide direct marginal interpretation of the regression parameters, by proposing the so-called marginalized multilevel model (Heagerty, 1999; Heagerty and Zeger, 2000). Emphasis is still on time-to-event data. Apart from model formulation, estimation methods (both maximum likelihood and pairwise likelihood) are discussed. A flexible joint multilevel modeling framework for repeated, time-to-event data is discussed in Chapter 6, where subject-specific interpretations of the parameters are presented at first (formulated conditionally upon the random

effects) and later on, extended to population-averaged interpretations of the parameters by using the so-called marginalized multilevel model (MMM) from Chapter 5. The joint model is aimed at modeling two outcomes that occur simultaneously as well as recognizing the relationship between the two outcomes. In this thesis, attention is given to the so-called shared-parameter framework (SPM) for random effects (Tsiatis and Davidian, 2004; Verbeke et al, 2010; Rizopoulos, 2008, 2011, 2012a), which models the two outcomes separately, conditionally upon the random effect, and merge them together with correlated random effects (representing the relationship among the two outcomes). Specific cases of this well thought-through framework are established in the repeated, overdispersed time-to-event setting, where at least one outcome is modeled with the proposed hierarchical/marginal combined model from Chapters 4 and 5.

A common problem that often arises in research is that of missing data. This has led to the development of selection, pattern-mixture, and shared-parameter missing data modelling frameworks (Molenberghs and Kenward, 2007). These frameworks have further been supplemented with characterizations of missing value mechanisms (Rubin, 1976; Molenberghs et al, 1998; Creemers et al, 2011), under the missing completely at random (MCAR), missing at random (MAR), and the missing not at random (MNAR) taxonomy. Given that models for missing data often make unverifiable assumptions about the missing value mechanism, a recurring theme is that of sensitivity analysis (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005; Creemers et al, 2010). As assumptions regarding the missing value mechanism are varied, the stability of inferences, or lack thereof, provides a guide on the caution with which the inferences need to be embraced. Undeniably, there is a strong connection between the missing data setting in a longitudinal context, on the one hand, and the joint longitudinal and time-to-event setting, on the other. Conceptually, the two settings actually correspond, but with an added layer of complexity in the latter setting (Njagi et al., 2013c). The additional complexity stems from the fact that data can now be coarsened in various ways: the longitudinal sequence can be incomplete; the time-to-event outcome can be censored; both of these can occur simultaneously. Coarsening refers to the phenomenon that data observed are less refined than the, possibly counterfactual, full data. This conceptual correspondence will be the focus of Chapter 7. We will take a slightly different perspective on joint models than is prevalent in the literature, and argue that conceptually, the two settings actually correspond. Based on this, we will build an extended shared random-effects survival-longitudinal joint model, similar in spirit to that of Creemers et al (2011) in the context of longitudinal data subject to missing observations, but now transposed to the current more complex setting. Within the extended framework, we will provide a characterization of MAR, consistent to the one in the missing data setting. We will then provide some reflections on the complexity of model formulation in the extended setting. The extended random effects structure will then be utilized for sensitivity analysis.

Routinely, after formulating and fitting a model, an assessment of the model fit and a diagnostic analysis is advisable. In Chapter 8, the assessment of the distribution of the random effects is argued in the proposed CM. Efendi et al (2013) provided a goodness-of-fit test based on the gradient function, which will be discussed in detail. Additionally, an appropriate test statistic is formulated, where, under the null-hypothesis, it is approximated using bootstrap methods. Chapter 9, on the other hand, covers a local influence diagnostics for the detection of influential subjects in the CM (Rakhmawati et al, 2014).

At first, background of the general theory behind local influence is handled, while afterwards, this is extended to the proposed CM.

Chapter 10 focuses on the problems of enriched data and the occurrence of non-identifiability. In the former one, data may be coarsened and/or augmented, two different phenomena which, together, constitute enriched data. Coarsened is that the actually observed data is less detailed than what is planned (e.g. incomplete data, censoring survival data settings, etc.), and augmented when the observed data are hypothetically but conventionally supplemented with structures such as random effects. Fitting models for enriched data combines evidence arising from empirical data with non-verifiable model components, i.e., that are purely assumption driven. Potentially dangers can arise from this, in the sense that an entire class of models can be assigned to any given model, with all of its members producing the same fit to the observed data but arbitrary regarding the unobservable parts of the enriched data. Non-identified parts can be replaced arbitrarily, without altering the fit to the observed data but with potentially non-trivial consequences for inferences and substantive conclusions (Verbeke and Molenberghs, 2010; Molenberghs et al, 2012). Therefore, assumptions should be supported by substantive considerations or be made part of a sensitivity analysis (Molenberghs et al, 2012), indicating that acceptable goodness-of-fit to the observed data (cf. Chapter 8) cannot be used as the sole justification for the analysis.

Finally, after applying the theoretically discussed framework in practice (Chapter 11), a global conclusion and outlook of this thesis will be given in Chapter 12.

# Chapter 2

# Motivating Case Study

Since this thesis mainly focuses on time-to-event data, a prevention trial, where children who are at a high risk of developing asthma are involved, is considered. A brief discussion of this dataset is considered in Section 2.1, while analysis are done in Chapter 11.

## 2.1 Recurrent Asthma Attacks in Children

The asthma data have been studied in Duchateau and Janssen (2008), and take the form of repeated time-to-event outcomes. Asthma is occurring more and more frequently in very young children, i.e., between 6 and 24 months. Therefore, a new application of an existing anti-allergic drug is administered to children who are at higher risk for developing asthma in order to prevent it. A prevention trial is set up with such children randomized to placebo or experimental drug, and the asthma events that developed over time are recorded in a diary. Typically, a patient has more than one asthma event. The intermittent events are thus clustered within a patient and ordered in time. The data are presented in a calendar time format, where the time at risk for a particular event is the time from the end of the previous event (asthma attack) to the start of the next event (start of the next asthma attack). A patient has different periods at risk throughout follow-up, which are separated either by an asthmaic event that lasts one or more days, or by a period in which the patient was not under observation. The start and end dates of each such risk period are required, together with the status indicator to denote whether the end of the risk period corresponds to an asthma attack or not. Data for the first patient are listed in Table 2.1.

**Table 2.1:** *Asthma data. The first four data points for the first two children, where column labeled 'Status' refers to whether* (1) *or not* (0) *censoring occurred.*

| Patient ID | Begin | End | Status | Drug |
|---|---|---|---|---|
| 1 | 0 | 15 | 1 | 0 |
| 1 | 22 | 90 | 1 | 0 |
| 1 | 96 | 325 | 1 | 0 |
| 1 | 329 | 332 | 1 | 0 |
| 2 | 0 | 180 | 1 | 1 |
| 2 | 189 | 267 | 1 | 1 |
| 2 | 273 | 581 | 1 | 1 |
| 2 | 582 | 600 | 0 | 1 |

# Chapter 3

# Overview of the Key Ingredients

In this chapter, an overview is provided of the key ingredients in the development of our proposed CM. These ingredients are a useful general tool for fully understanding the developed modeling framework (Molenberghs et al, 2010), and all its corresponding special cases, such as dealing with overdispersion and to account for data hierarchies, such as longitudinal (repeated over time) data, separately, without changing much in the terminology of the framework.

In Section 3.1, a description is given of the conventional exponential family and generalized linear modeling (GLM) (Nelder and Wedderburn 1972, Agresti 2002), which serves as basic building block in the development of the CM. Section 3.2 deals with models to accommodate overdispersion (Hinde and Demétrio 1998ab). Verbeke et al (2000) and Molenberghs et al (2005) introduced normal random effects to address dependency in Gaussian and non-Gaussian outcomes, respectively, resulting in the so-called linear mixed model (LMM) and generalized linear mixed model (GLMM). These modeling techniques will be handled in Section 3.3, and increased importance is given to the latter one. Section 3.4 (main interest of this chapter) focuses on the combined modeling framework (introduced by Molenberghs et al, 2000), incorporating a conjugate random effect and normal random effect to simultaneously address correlation within repeated measures sequences and overdispersion.

This chapter outlines a general framework for the CM, while specific attention is given to the time-to-event setting in the next chapter (Chapter 4).

## 3.1 Standard Generalized Linear Models

The occurency of (non-)Gaussian outcomes is often encountered in many practical applications. For example, Huang et al (1983) used a non-Gaussian statistical model for surface elevation of nonlinear random wave fields in the world of physics, while Jansen (2010) analyzed, in the field of medical sciences, diffusion-weighted MR Imaging in head and neck squamous cell carcinoma with a non-Gaussian model. Analyzing these outcomes often happen with univariate generalized linear models (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; J$\phi$rgensen, 1987), based on the so-called exponential family distributions.

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

$$f(y) \equiv f(y \mid \eta, \phi) = \exp\{\phi^{-1} \cdot [y \cdot \eta - \psi(\eta)] + c(y, \phi)\}, \tag{3.1}$$

for a specific set of of unknown parameters $\eta$ ("natural parameter") and $\phi$ ("dispersion parameter"), and for known functions $\psi(\cdot)$ and $c(\cdot, \cdot)$.

After some straightforward derivations (Appendix A.1), it can easily be shown that

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

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

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

The exponential family embraces a lot of distributions. The best known are the normal (for continuous data), binomial/Bernoulli (for binary data), Poisson (for count data) and exponential (for time-to-event data) distribution. These distributions, along with their exponential family elements, are extensively described in Molenberghs et al (2010). In this thesis, emphasis is placed on the Weibull and Exponential distribution for modeling time-to-event data. In the exponential case, one assumes

$$f(y) = \varphi \cdot e^{-\varphi \cdot y}, \tag{3.4}$$

with mean $\varphi^{-1}$ and variance $\varphi^{-2}$. In the Weibull case, this extends to

$$f(y) = \varphi \cdot \rho \cdot y^{\rho - 1} \cdot e^{-\varphi \cdot y^{\rho}}, \tag{3.5}$$

where the mean and variance are expressed by $\varphi^{-1/\rho} \cdot \Gamma(\rho^{-1} + 1)$ and $\varphi^{-2/\rho} \cdot [\Gamma(2 \cdot \rho^{-1} + 1) - \Gamma(\rho^{-1} + 1)^2]$, respectively.

A few comments are in place. First, the Weibull model does not belong to the exponential family in a conventional sense, unless in a somewhat contrived fashion where $Y$ is replaced by $Y^{\rho}$. Second, setting $\rho = 1$ in the Weibull case leads to the exponential time-to-event distribution. Therefore, special attention is given to the more general Weibull distribution. Third, the mean $\mu$ (through the function $\eta$) can depend on $p$-dimensional vectors of covariate values $\mathbf{x}_i$ for outcome $Y_i$, with $i = 1, \ldots, N$. More precisely, $\mu_i = h(\eta_i) = h(\mathbf{x}_i \cdot \boldsymbol{\xi})$, for a known function $h(\cdot)$ (= the inverse link function) and a vector of $p$ fixed, unknown regression coefficients $\boldsymbol{\xi}$. The model is termed 'generalized linear model '(GLM), where the link function $h(\cdot) = \psi'(\cdot)$ is called the natural link function, in which case $\eta_i = \mathbf{x}_i \cdot \boldsymbol{\xi}$. For the Weibull and exponential model, the decomposition $\varphi = \lambda \cdot e^{\mu}$ is often employed. In this situation, while $\mu$ is a component of the mean function, it is in itself not equal to the mean. Fourth, $\Gamma(\cdot)$ represents the gamma function in the mean and variance expressions. Fifth, maximum likelihood or quasi-likelihood can be used for parameter estimation. When quasi-likelihood methods are employed (McCullagh and Nelder, 1989; Wedderburn, 1974; Molenberghs and Verbeke, 2005), no full distributional assumptions are made, but one rather restricts to specifying the first two moments (3.2) and (3.3). In such an instance, the variance function $v(\mu)$ can be chosen in accordance with a particular member of the exponential family. If not, then parameters cannot be estimated using maximum likelihood principles. Instead, a set of estimating equations needs to be specified, the solution of which is referred to as the quasi-likelihood estimates.

## 3.2 Overdispersion Models

One of the key features of the GLM framework and many of the exponential family members is the so-called mean-variance relationship, where the variance is a deterministic function ($v(\cdot)$) of the mean. However, in many practical situations (for count and time-to-event data, for example), this restriction is not in line with a particular set of data, and may cause serious flaws in point and precision estimation and inference on important parameters (Paul and Plackett, 1978; Cox, 1983; Breslow, 1990). This may lead to incorrect conclusions; for instance, a treatment which does not have a significant effect could be assessed as if it had an effect. Two phenomena can occur: overdispersion and underdispersion. The former one arises when the observed variance from the data is greater than the theoretical variance (restricted by the mean-variance relationship) from the model, while the latter one is obtained when the observed variance is smaller than the theoretical variance. In this thesis, emphasis is placed on overdispersion. For example, Mwangi et al (2008) provide evidence for overdispersion in the distribution of clinical malaria episodes in children.

To handle the problem of overdispersion, a number of extensions have been proposed. Hinde and Demétrio (1998a, 1998b) provide general treatments of overdispersion, by modifying the GLM. The most popular models to encounter overdispersion in many practical situations are the negative-binomial model (Breslow, 1984; Lawless, 1987; Gardner et al, 1995; Engel, 1984; Manton et al, 1981) and zero-inflated Poisson model (Lambert, 1992; Hall, 2000) for count data and the beta-binomial model (Wilcox, 1981; Skellam, 1948) and Bahadur model (Bahadur, 1961; Molenberghs and Verbeke, 2005) for categorical data. Indeed, Mwangi et al (2008) concluded that the pattern of clinical malaria episodes follows a negative binomial distribution.

A natural and straightforward step is to allow the overdispersion parameter $\phi \neq 1$, so that (3.3) produces $\text{Var}(Y) = \phi \cdot v(\mu)$. This is in line with the so-called moment-based approach, but can also be engendered by fully parametric assumptions. Another, more convenient route is through the so-called two-stage approach, i.e., by placing a distribution on the model parameter (also known as a random effect). The principle behind this approach is to consider a distribution for the outcome of interest, given the random effect $f(y_i \mid \theta_i)$ which, combined with a model for the random effect $f(\theta_i)$, produces the marginal model:

$$f(y_i) = \int_\theta f(y_i \mid \theta_i) f(\theta_i) d\theta_i. \tag{3.6}$$

Until now, a general discussion was obtained for the latter approach. In case of time-to-event data, where the Weibull and exponential distribution seems appropriate choices for the hierarchical model, several extensions are possible. Often, preference is given to Gamma conjugate random effects, also known as frailties (Duchateau en Janssen, 2008), giving rise to the exponential-gamma and Weibull-gamma models. The model elements are listed in Table 3.1. Choosing the Gamma distribution (with parameters $\alpha$ and $\beta$) has the advantage of (1) satisfying the mean's scale for time-to-event outcomes (Section 4.1) and (2) obtaining closed forms for the marginal mean and variance, and even for the entire marginal distribution (Molenberghs et al, 2010). The latter one arises from the concept of conjugacy (Cox and Hinkley, 1974; Lee, Nelder, and Pawitan, 2006; Agresti, 2002).

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. To be more precise, mathematically, it can be said that the hierarchical and random effects densities are conjugate if and only if they can be written in the generic forms

$$f(y \mid \theta) = \exp\{\phi^{-1} \cdot [y \cdot h(\theta) - g(\theta)] + c(y, \phi)\}, \tag{3.7}$$

$$f(\theta) = \exp\{\gamma \cdot [\psi \cdot h(\theta) - g(\theta)] + c^*(y, \psi)\}, \tag{3.8}$$

where $g(\theta)$ and $h(\theta)$ are functions, $\phi$, $\gamma$ and $\psi$ are parameters, and the additional functions $c(y, \phi)$ and $c^*(y, \psi)$ are so-called normalizing constants. By using the two-stage approach (3.6) with (3.7) and (3.8), it can be shown that the marginal model equals

$$f(y) = \exp\left[c(y, \phi) + c^*(y, \psi) + c^*\left(\phi^{-1} + \gamma, \frac{\phi^{-1} \cdot y + \gamma \cdot \psi}{\phi^{-1} + \gamma}\right)\right], \tag{3.9}$$

**Table 3.1:** *Model elements for the Weibull-gamma and exponential-gamma models*

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

If the time-to-event data corresponds to a hierarchical structure, with $Y_{ij}$ denoting the $j$th time-to-event outcome measured for cluster (subject) $i$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and $\mathbf{Y}_i$ the $n_i$-dimensional vector of all measurements available for cluster $i$, then the scalar $\theta_i$ becomes a vector $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{in_i})$, with $\mathrm{E}(\boldsymbol{\theta}_i) = \boldsymbol{\mu}_i$ and $\mathrm{Var}(\boldsymbol{\theta}_i) = \Sigma_i$. In line with the univariate case, the model produces $\mathrm{E}(\mathbf{Y}_i) = \boldsymbol{\mu}_i$ and $\mathrm{Var}(\mathbf{Y}_i) = M_i + \Sigma_i$, where $M_i$ is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the diagonal. Note that a diagonal structure of $M_i$ reflects the conditional independence assumption: all dependence between measurements on the same unit stems from the random effects. Generally, a versatile class of models results. For example, assuming that the components of $\boldsymbol{\theta}_i$ are independent, a pure overdispersion model follows, without correlation between the repeated measures. On the other hand, assuming $\theta_{ij} = \theta_i$, that is, that all components are equal, then $\mathrm{Var}(\mathbf{Y}_i) = M_i + \sigma_i^2 \cdot J_{n_i}$, where $J_{n_i}$ refers to the ($n_i$ x $n_i$)-dimensional unit

matrix in the sense that all elements equal one. Such a structure can be seen as a general version of compound symmetry. Of course, one can also combine general correlation structures between the components of $\theta_i$. Additionally, the proposed Weibull-gamma and exponential-gamma model are derived for a two-parameter gamma density. In a gamma frailty context (Duchateau and Janssen, 2007), it is customary to set $\alpha_j \cdot \beta_j = 1$ (assuming different parameters for each outcome $j$), for reasons of identifiability.

Until now, no attention has been given to the possible structures of the Gamma (conjugate) random effects in the Weibull-gamma and exponential-gamma model. Assuming a scalar vector $\boldsymbol{\theta}_i$ (Molenberghs et al, 2007; Aregay et al, 2013a), where the extra variation is assumed constant over time, does not always provide the best choice and/or even may not be satisfied in some cases. In addition, the dispersion parameter may be different across groups. For example, in an experiment on salamander matings (McCullagh and Nelder, 1989), Lin (1997) has studied whether or not there is heterogeneity in the mating success probability among male and female salamanders, using a generalized linear model with random effects. To attribute to the model more flexibility in terms of encompassing multiple sources of variation, Aregay et al (2013c) extended the (combined) model to so-called stratified overdispersion models, in terms of a Bayesian framework (Section 3.5.2). The rest of this section is dedicated on this extended framework.

In case of the Weibull-gamma and exponential-gamma model, the prior distribution of the overdispersion parameter is specified

$$\theta_{ij} \sim \text{Gamma}(\alpha, \beta), \tag{3.10}$$

Here, the distribution of the overdispersion parameter is assumed to be the same across all timepoints and/or other covariates in the model. Such an approach was followed by Molenberghs et al (2007, 2010) within the frequentist framework, and by Aregay et al (2013ab) within the hierarchical Bayesian framework. Aregay et al (2013c) formulated a generalized model for $\theta_{ij}$ that allows the distribution of the overdispersion parameter to vary across covariate levels, by specifying the dependency of the Gamma distribution on the covariates via a model for the hyper-parameters. Similar to a GLM (Section 3.1), the generalized model for $\theta_{ij}$ in the Weibull-gamma and exponential-gamma case has three components: (1) a Gamma prior distribution of $\theta_{ij}$, (2) a link function $g(\cdot)$, and (3) a linear predictor, used to model the dependency of the Gamma distribution on covariates via parametrization of the hyper-parameters.

To stay in line with the followed terminology of Aregay et al (2013c), $\mathbf{X}_{OD}$ denotes a known design matrix and $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ a parameter vector. The generalized model for $\theta_{ij}$ is then formulated by

$$\boldsymbol{\theta}_i \sim \text{Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{3.11}$$

$$g(\boldsymbol{\alpha}) = \mathbf{X}_{OD} \cdot \boldsymbol{\tau}_1, \tag{3.12}$$

$$g(\boldsymbol{\beta}) = \mathbf{X}_{OD} \cdot \boldsymbol{\tau}_2, \tag{3.13}$$

where $g(\cdot)$ is a chosen link function and $\mathbf{X}_{OD} \cdot \boldsymbol{\tau}_1$, respectively $\mathbf{X}_{OD} \cdot \boldsymbol{\tau}_2$, are the linear predictors for the hyper-parameters of $\boldsymbol{\alpha}$, respectively $\boldsymbol{\beta}$, in Gamma$(\boldsymbol{\alpha}, \boldsymbol{\beta})$. For reasons of identifiability, $\alpha_j \cdot \beta_j = 1$ is assumed. Many different choices can be made for $\mathbf{X}_{OD}$ and $\boldsymbol{\tau}$, depending on the research question of interest. Here, a few basic choices are considered.

1. *An unstructured time dependent Gamma distribution for the overdispersion par. $\theta_{ij}$*

The aim is to model the dependency of the overdispersion parameter on time, by defining time specific hyper-parameters $\boldsymbol{\tau}_1 = (\alpha_1, \alpha_2, \ldots, \alpha_T)^{'}$, $\boldsymbol{\tau}_2 = (\beta_1, \beta_2, \ldots, \beta_T)^{'}$ and the design matrix $\mathbf{X}_{OD}$ specifies a ($T$ x $T$)-dimensional identity matrix ($T$ is the number of time points). Thus, the distribution of $\theta_{ij}$ is time dependent, $\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j)$ ($j = 1, 2, \ldots, T$). Mathematically, for $j = 1, 2, \ldots, T$,

$$\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j), \tag{3.14}$$

$$(g(\alpha_1), g(\alpha_2), \cdots, g(\alpha_T)) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \cdot (\alpha_1, \alpha_2, \ldots, \alpha_T)^{'}, \tag{3.15}$$

$$(g(\beta_1), g(\beta_2), \ldots, g(\beta_T)) = \left[ g\left(\frac{1}{\alpha_1}\right), g\left(\frac{1}{\alpha_2}\right), \ldots, g\left(\frac{1}{\alpha_T}\right) \right]. \tag{3.16}$$

This choice describes the dependency of the Gamma distribution of $\theta_{ij}$ on time by specifying different (across time) prior distributions for $\theta_{ij}$ (i.e., different hyper-parameters in the Gamma distribution), and may cause overparameterization. Therefore, careful consideration needs to be obtained for choosing this approach. Sensitivity analysis (Section 10.2) can be an appropriate tool here.

2. *A linear time dependent Gamma distribution for the overdispersion parameter $\theta_{ij}$*

In this case, a linear dependency between $\alpha_j$ and time is used, by defining the design matrix $\mathbf{X}_{OD}$ and hyper-parameter vector $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, respectively,

$$\mathbf{X}_{OD} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_T \end{pmatrix}, \boldsymbol{\tau}_1 = (\alpha_0, \alpha_1)^{'}, \text{ and } \boldsymbol{\tau}_2 = (\beta_0, \beta_1)^{'}.$$

It has the advantages of (1) avoiding overparametrization and (2) having an easy evaluation whether or not a linear time dependent overdispersion model is appropriate for the given dataset by checking the 95% confidence interval for $\alpha_1$ and/or $\beta_1$. Mathematically, formula (3.14) is kept, with $g(\boldsymbol{\alpha}) = \mathbf{X}_{OD} \cdot (\alpha_0, \alpha_1)^{'}$ or $g(\alpha_j) = \alpha_0 + \alpha_1 \cdot t_j$ in the first expression and $g(\boldsymbol{\beta}) = g(\boldsymbol{\alpha}^{-1})$ or $g(\beta_j) = g\left(\frac{1}{\alpha_j}\right)$ in the second one ($j = 1, 2, \ldots, T$).

3. *A covariate dependent Gamma distribution for the overdispersion parameter $\theta_{ij}$*

In many statistical areas, attention is given to whether or not covariates produce a significant effect on the outcome. In this context, we allow GLM (3.11)–(3.13) for inclusion of other covariates that might influence the Gamma distribution of the overdispersion parameters. Often, especially in the medical area, researchers

are interested in whether or not a treatment benefits significantly from a placebo for patients $i$ ($i = 1, 2, \ldots, N$), which were followed over time with time points $j$ ($j = 1, 2, \ldots, n_i$). Modeling this into the GLM (3.11)-(3.13), a binary covariate $I_i$ (representing the treatment indication for a specific patient) is considered:

$$I_i = \begin{cases} 1 & \text{if subject } i \text{ received the treatment} \\ 0 & \text{otherwise} \end{cases}.$$

Supposing that the prior distribution of $\theta_{ij}$, i.e. (3.11), is now formulated by $\theta_{ij} \sim \text{Gamma}(\alpha_i, \beta_i)$. Additionally, the dimension of $\boldsymbol{\alpha}$ is not the number of time points anymore but the number of treatment groups. Using these constraints, the model takes the form:

$$\theta_{ij} \sim \text{Gamma}(\alpha_i, \beta_i), \tag{3.17}$$

$$(g(\alpha_1), g(\alpha_2), \cdots, g(\alpha_N)) = \begin{pmatrix} I_1 & 1 - I_1 \\ I_2 & 1 - I_2 \\ \vdots & \vdots \\ I_N & 1 - I_N \end{pmatrix} \cdot (\alpha_1, \alpha_2)', \tag{3.18}$$

$$(g(\beta_1), g(\beta_2), \ldots, g(\beta_N)) = \left[ \left( \frac{1}{\alpha_1} \right), g\left( \frac{1}{\alpha_2} \right), \ldots, g\left( \frac{1}{\alpha_N} \right) \right]. \tag{3.19}$$

4. *A constant Gamma distribution for the overdispersion parameter $\theta_{ij}$*

To conclude this discussion, an overview is given on the most easy choice (Molenberghs et al, 2007; Aregay et al, 2013a). Here, the distribution of $\theta_{ij}$ is the same across all time points, by specifying a ($n_i$ x $n_i$)-dimensional identity matrix for the design matrix and constant hyper-parameter $\boldsymbol{\tau}_1 = (\alpha, \alpha, \ldots, \alpha)'$. Hence, formula (3.14) holds with $g(\boldsymbol{\alpha}) = (\alpha, \alpha, \ldots, \alpha)$ and $g(\boldsymbol{\beta}) = g(\boldsymbol{\alpha}^{-1})$.

## 3.3 Models with Normal Random Effects

Alternatively, it is possible to include normal random effects in the linear predictor $\eta_i$ of the generalized linear model. In Section 3.1, this predictor is set equally to $\mathbf{x}_i \cdot \boldsymbol{\xi}$, i.e., a general assumption that is often made in statistical analysis. Extending this with random effects gives rise to the family of known as generalized linear mixed model (Thall and Vail, 1990; Dean, 1991; Engel and Keen, 1994; Wolfinger and O'Connell, 1993; Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005). It has the great advantages of (1) taking into account the hierarchical and repeated (e.g. longitudinal) data structures, by a single modeling framework and (2) exploring within-subject and between-subject correlations from the data. Analyzing datasets with standard GLM models does not have these advantages. For example, if a standard Poisson regression model is used to analyze the number of epileptic seizures for hundred subjects, hundred Poisson regression models are modeled separatly, without having the opportunity to explore between- and within-subject correlations. By applying the Poisson linear mixed model, one model is used to analyze this outcome, where the between- and within-subject correlations are taken into

account by the random effects. Often, empirical Bayes estimation (Laird and Ware, 1982; Schall, 1991) is used to estimate these correlations.

First, a discussion of the linear mixed model (LMM) is given, used for continuous repeated measures, followed by the generalized linear mixed model (GLMM), which is the most popular method for non-Gaussian repeated measures (e.g., count, binary, time-to-event, etc.). The latter one is not only a relatively straightforward extension of the GLM for independent data (Section 3.1) to the context of hierarchically organized data, on the one hand, and the linear mixed model (Verbeke and Molenberghs, 2000), on the other hand, but there is also a wide range of software tools available for fitting such models (e.g. Broström, 2003; Littell, 2006; Faraway, 2006).

### 3.3.1 Linear Mixed Models

The LMM has become an important tool for the analysis of continuous longitudinal and/or hierarchical data, by the inclusion of additional random effects in the linear predictor. The distribution of these random effects are often assumed normal, due to conjugacy reasons (3.7)–(3.8). In this thesis, normal random effects are assumed in the linear predictor $\eta_i$, but keeping in mind that other possible choices can be made. Verbeke and Molenberghs (2000) devoted an entire text to LMM, along with its extensions. A wide variety of software packages are available for fitting the model, such as the SAS procedure MIXED (Wolfinger and Chang, 1998; Littell et al, 2006), the SPlus function lme (Pinheiro and Bates, 2000), the MLwiN package (Lawson et al, 2003; Rasbash et al, 2014), etc.

The LMM assumes that the $n_i$-dimensional vector of repeated outcome measurements $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ for cluster (subject) $i$, $i = 1, \ldots, N$, follow a linear regression model. The fixed regression parameters $\boldsymbol{\xi}$ represent a population-specific interpretation (i.e., the same for all clusters), while the remaining parameters, i.e., the random effects $\mathbf{b}_i$, are used for between-cluster interpretations. The LMM framework can generally be expressed by

$$\mathbf{Y}_i = \mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{Z}_i \cdot \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{3.20}$$

$$\begin{bmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim N\left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \Sigma_i \end{bmatrix} \right), \tag{3.21}$$

where $\boldsymbol{\xi}$ is a ($p$ x 1)-dimensional vector of unknown fixed parameters, $\mathbf{b}_i$ is a ($q$ x 1)-dimensional vector of random effects, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are known design matrices with dimensions of respectively $n_i$ x $p$ and $n_i$ x $q$, $\boldsymbol{\varepsilon}_i$ is a ($n_i$ x 1)-dimensional vector of unobserved measurement errors. Similar as before, $N$ denotes the number of clusters (subjects), while $n_i$ represents the number of repeated measurements for cluster (subject) $i$. Additionally, the random effects $\mathbf{b}_i$ and random error term $\boldsymbol{\varepsilon}_i$ are assumed to be independent from each other, where $\mathbf{b}_i$ follows a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $D$ and $\boldsymbol{\varepsilon}_i$ follows a normally distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\Sigma_i$.

Equivalently, the LMM formulation (3.20)–(3.21) can be re-expressed in a clear hierarchi-

cal notation of the form

$$\mathbf{Y}_i \mid \mathbf{b}_i \sim N(\mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{Z}_i \cdot \mathbf{b}_i, \Sigma_i), \tag{3.22}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \tag{3.23}$$

Due to conjugacy reasons (3.7)–(3.8), a closed form is adjusted for the marginal model (Verbeke and Molenberghs, 2000):

$$\mathbf{Y}_i \sim N(\mathbf{X}_i \cdot \boldsymbol{\xi}, \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} + \Sigma_i), \tag{3.24}$$

Though the marginal formulation (3.24) naturally follows from the hierarchical formulation (3.22)–(3.23), both models are not equivalent. For an elaborate discussion about the differences, reference is given to Verbeke and Molenberghs (2000). On top of this, several structures are possible for $\mathbf{Z}_i$ and $\mathbf{b}_i$, depending on the research question. Two of the most frequently used models are the so-called random-intercept and random-slope models, often employed in longitudinal structures (Zuur et al, 2007; Zeger and Liang, 1992).

1. *The random-intercept model for Gaussian outcomes in a longitudinal, hierarchical data structure*

   For simplicity, only two fixed effects $\xi_0$ and $\xi_1$ are assumed here. The former one describes the intercept, while the latter one represents the time-dependent coefficient. Furthermore, the same assumptions are used as in (3.20)–(3.21). The hierarchical random-intercept model for Gaussian outcomes in a longitudinal, hierarchical data structure takes the form

$$\mathbf{Y_i} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \cdot \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \cdot b_{0i} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, \tag{3.25}$$

$$\begin{bmatrix} b_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} d & \begin{pmatrix} 0 & \cdots & 0 \end{pmatrix} \\ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} & \Sigma_i \end{bmatrix} \right). \tag{3.26}$$

   This is a specific case of the general LMM (3.20)–(3.21), in the sense that $\mathbf{b}_i = b_{0i}$, i.e., only one random effect $b_{0i}$ is used to capture the between-subject variability. Here, the evolution over time stays constant for all subjects $i$.

2. *The random-slope model for Gaussian outcomes in a longitudinal, hierarchical data structure*

   Extending the random-intercept model (3.25)–(3.26) with an extra time-dependent random effect $b_{1i}$, gives rise to the so-called random-slope model. This model is

formulated by

$$\mathbf{Y_i} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \cdot \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} + \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \cdot \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, \tag{3.27}$$

$$\begin{bmatrix} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim N \left( \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} D & \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} & \Sigma_i \end{bmatrix} \right). \tag{3.28}$$

In this case, two random effects $b_{0i}$ and $b_{1i}$ are used for subject $i$. The former one captures the between-subject variability, similar to the random-intercept model, while the latter one is used to capture within-subject variability.

### 3.3.2 Generalized Linear Mixed Models

The LMM framework (for continuous clustered, repeated measures, and longitudinal studies, for example, or whenever there are data hierarchies) can easily be extended to the standard GLM framework (Section 3.1), giving rise to the so-called GLMM (Engel and Keen, 1994; Breslow and Clayton, 1993, Wolfinger and O'Connell, 1993) for non-Gaussian repeated measurements within a hierarchical data structure. Similar to LMM, it accommodates the correlation between repeated measurements and hierarchical structure by means of normal random effects in the linear predictor and to some extent for overdispersion as well, but uses the exponential family distribution (3.1) instead of the normal distribution. The same terminology is adopted from Section 3.1 and 3.3.

Assume that, conditionally upon the $q$-dimensional random effects $\mathbf{b}_i \sim N(\mathbf{0}, D)$, the outcomes $Y_{ij}$ are independent with exponential-family densities of the form

$$f_i(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \phi) = \exp\{\phi^{-1} \cdot [y_{ij} \cdot \lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\}, \tag{3.29}$$

with

$$h^{-1}[\psi'(\lambda_{ij})] = h^{-1}(\mu_{ij}) = h^{-1}[\mathrm{E}(Y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi})] = \eta_{ij} = \mathbf{x}_{ij}' \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}' \cdot \mathbf{b}_i, \tag{3.30}$$

where $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ are $p$- and $q$-dimensional vectors of known covariate values, $\lambda_{ij}$ represents the natural parameter and $\eta_{ij}$ the linear predictor with random effect $\mathbf{b}_i$. Let $f(\mathbf{b}_i \mid D)$ be the density of the $N(\mathbf{0}, D)$ for the random effects $\mathbf{b}_i$. Often, unlike the LMM, no closed forms are obtained for integral (3.6), nor for the corresponding moments $\mathrm{E}(Y_{ij}^k)$ $(k = 1, 2, 3, \dots)$. When making inferences on these marginal models, a suite of computational techniques, e.g., using Taylor series expansions (Hildebrand, 1962) and numerical integration (Davis and Rabinowitz, 2007), has been derived to approximate the marginal likelihood numerically (Molenberghs and Verbeke, 2005; Raudenbush et al, 2000; Wolfinger and Lin, 1997). The marginal likelihood function is given by

$$L(\boldsymbol{\vartheta}, D) = \prod_{i=1}^{N} \left[ \int_b \left( \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \boldsymbol{\vartheta}, \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \right) \cdot d\mathbf{b}_i \right]. \tag{3.31}$$

Here, $\boldsymbol{\vartheta}$ groups all parameters in the conditional model for $\mathbf{Y}_i$, given the random effects. Alternatively, other estimation methods such as pairwise likelihood and Bayesian estimation can be used. These estimation approaches are covered in Section 4 (for the proposed hierarchical Weibull-gamma-normal model) and 5 (for the marginal Weibull-gamma-normal model), for time-to-event data. Furthermore, local influence is described from the marginal likelihood (3.27) for the specific Weibull-normal model (Section F.2). It should be noted that the GLMM framework is a bit less common for survival data, where so-called frailty models (Duchateau and Janssen, 2007), rather of the type with conjugate random effects, are more standard.

## 3.4 Flexible Modeling Framework to Combine Overdispersion and Normal Random-Effects

Although the GLMM from Section 3.3.2 accommodates the correlation between repeated measures in a hierarchical data structure, even for some of the overdispersion as well, limitations can arise. Booth et al (2003) and Molenberghs et al (2007) studied this model and concluded that the formulated GLMM (3.29)-(3.30) often inadequately fits the data in an overdispersed, repeated structure, where overdispersion and correlation between repeated measurements can occur simultaneously. To deal with this issue, Molenberghs et al (2007) extended the GLMM to a so-called combined model (CM), by including an extra random effect. This random effect enters the mean directly as a multiplicative factor, implying that mean scaling must be satisfied. For example, for time-to-event outcomes, the inserted random effect must have support over the positive half line. Furthermore, Molenberghs et al (2010) generalized this CM into the exponential family framework. This elegant proposition is adopted, and will be specified and declared for time-to-event data (Section 4 and 5).

Adding an extra random effect to the GLMM has the disadvantage of decreasing the chance of having conjugacy among the chosen hierarchical and random effect distributions. To resolve this issue, like mentioned in Section 3.3.2, generic approximations can be used to approximate marginal model elements (Molenberghs et al, 2010). These approximations for the CM are briefly discussed in Section 3.4.2, along with the principle of strong conjugacy in the CM framework (Molenberghs et al, 2010).

### 3.4.1 General Model Formulation

Assume $Y_{ij}$ to be the same as in Section 3.2. Combining both the overdispersion random effects $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{in_i})'$ and normal random effects $\mathbf{b}_i$ into the exponential family framework (3.1) led Molenberghs et al (2010) to the general CM family:

$$f_i(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \boldsymbol{\theta}_i, \phi) = \exp\{\phi^{-1} \cdot [y_{ij} \cdot \lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\}, \tag{3.32}$$

with conditional mean

$$\mu_{ij}^c = \mathrm{E}(Y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \theta_{ij}) = \theta_{ij} \cdot \kappa_{ij}, \tag{3.33}$$

where $\theta_{ij} \sim \Upsilon_{ij}(\varpi_{ij}, \sigma_{ij}^2)$ for some distribution $\Upsilon_{ij}$, with mean $\varpi_{ij}$ and variance $\sigma_{ij}^2$, and $\kappa_{ij} = h(\eta_{ij}) = h(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i)$. Similar as before, $\mathbf{b}_i$ is assumed to be normally distributed with mean vector $\mathbf{0}$ an variance-covariance matrix $D$.

A few comments are in place. First, for generality of notation, all aspects in the distribution of $\theta_{ij}$ are subscripted by $i$ and $j$, whereas one might choose all distributions to be common to a particular measurement occasion $j$ or even common over values of $i$ and $j$. Second, the two parameters $\eta_{ij}$ and $\lambda_{ij}$ refer to the linear predictor and/or the natural parameter. The basic difference is that $\lambda_{ij}$ encompasses the random variables $\theta_{ij}$, which captures overdispersion at mean scale, whereas $\eta_{ij}$ refers to the "GLMM part" only. Third, it is convenient, but not strictly necessary, to assume that the two sets of random effects, $\theta_{ij}$ and $\mathbf{b}_i$, are independent of each other. Fourth, regarding the components $\theta_{ij}$ of $\boldsymbol{\theta}_i$, three useful special cases result from assuming that: (1) they are independent; (2) they are correlated, implying that the collection of univariate distributions $\Upsilon_{ij}(\varpi_{ij}, \sigma_{ij}^2)$ needs to be replaced with a multivariate one; and (3) they are equal to each other, useful in applications with exchangeable outcomes $Y_{ij}$. Fifth, choosing a conjugate choice for $\theta_{ij}$, not only has the advantage of respecting the range for the mean, but also that closed forms for the marginal mean and variance, and even for the entire marginal distribution, are possible (Molenberghs et al, 2010). Sixth, the relationship between the conditional mean and natural parameter now is

$$\lambda_{ij} = g(\mu_{ij}^c) = g(\theta_{ij} \cdot \kappa_{ij}). \tag{3.34}$$

Note that the function $g(\cdot)$ transforms the product $\theta_{ij} \cdot \kappa_{ij}$, whereas the function $h(\cdot)$ transforms the $\kappa_{ij}$ only. For the marginal mean of the outcome $Y_{ij}$, with the use of iterated-expectation-base calculations, we have:

$$\mathrm{E}(Y_{ij}) = \mathrm{E}(\theta_{ij}) \cdot \mathrm{E}(\kappa_{ij}) = \mathrm{E}[g^{-1}(\lambda_{ij})]. \tag{3.35}$$

The marginal likelihood function equals

$$\begin{aligned} L(\boldsymbol{\vartheta}, D, \boldsymbol{\varpi}, \Sigma) &= \prod_{i=1}^{N} L_i(\mathbf{y}_i \mid \boldsymbol{\vartheta}, D, \boldsymbol{\varpi}_i, \Sigma_i) \\ &= \prod_{i=1}^{N} \left[ \int_{\theta,b} \left( \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) \cdot f(\mathbf{b}_i \mid D) \cdot f(\boldsymbol{\theta}_i \mid \boldsymbol{\varpi}_i, \Sigma_i) \cdot d\mathbf{b}_i \cdot d\boldsymbol{\theta}_i \right) \right], \end{aligned} \tag{3.36}$$

where $L_i(\mathbf{y}_i \mid \boldsymbol{\vartheta}, D, \boldsymbol{\varpi}_i, \Sigma_i)$ presents the likelihood contribution of subject $i$ and $\boldsymbol{\vartheta}$ groups all parameters in the conditional model for $\mathbf{Y}_i$.

### 3.4.2 Generic Approximations for Marginal Model Elements

Even though formula (3.35) allows for explicit expressions of the marginal means in a good number of cases (Molenberghs et al, 2007; Molenberghs et al, 2010), this is not generally true for all cases. For example, Molenberghs et al (2007, 2010) derived explicit expressions for the means, variances, and marginal densities in a number of outcome types, such as normal, Poisson, and time-to-event. However, Molenberghs et al (2010)

explored that no closed forms for either the mean nor the variance follow in Bernoulli-type models for binary data with logit link and normal random effects $\mathbf{b}_i$, whether or not beta random effects $\theta_{ij}$ are present. Therefore, Molenberghs et al (2010) derived an approximate expression for the marginal model elements, using a Taylor series expansion (Hildebrand, 1962) around $\mathbf{b}_i = 0$:

$$\kappa_{ij} \approx h(\eta_{ij}) + h'(\eta_{ij}) \cdot \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i + \frac{1}{2} \cdot h''(\eta_{ij}) \cdot \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i \cdot \mathbf{b}_i^{'} \cdot \mathbf{z}_{ij}. \tag{3.37}$$

The generic marginal mean, variance and covariance approximations, and supplementary details, are provided in Appendix A.2.

### 3.4.3 Strong Conjugacy

Until now, the concept of conjugacy (Cox and Hinkley, 1974; Lee, Nelder, and Pawitan, 2006; Agresti, 2002) was discussed for only one random effect distribution at a time (Section 3.2). Extending the model with an extra random effect, results in a lower chance of having conjugacy among the hierarchical and random effects distribution. Therefore, it is of interest to explore under what conditions Model (3.32) still allows for conjugacy.

Molenberghs et al (2010) explored these conditions in the CM framework, where both normal and overdispersion random effects are included, and introduced the principle of strong conjugacy as a way of expressing in which cases conjugacy remains, even in the presence of normally distributed random effects $\mathbf{b}_i$. They considered conjugacy, conditional upon the normally-distributed random effect $\mathbf{b}_i$. To this effect, they wrote (suppressing nonessential arguments from the functions):

$$f(y \mid \kappa \cdot \theta) = \exp\{\phi^{-1} \cdot [y \cdot h(\kappa \cdot \theta) - g(\kappa \cdot \theta)] + c(y, \phi)\}, \tag{3.38}$$

generalizing (3.7), and retain (3.8). Here, the natural parameter is multiplicatively generalized to separate the transformed linear predictor $\kappa$ from the parameter $\theta$. Applying the transformation theorem to (3.37) leads to

$$f(\theta \mid \gamma, \psi) = \kappa \cdot f(\kappa \cdot \theta \mid \tilde{\gamma}, \tilde{\psi}), \tag{3.39}$$

Next, we request the parametric form (3.8) be maintained:

$$f(\kappa \cdot \theta) = \exp\{\gamma^* \cdot [\psi^* \cdot h(\kappa \cdot \theta) - g(\kappa \cdot \theta)] + c^{**}(\gamma^*, \psi^*)\}, \tag{3.40}$$

where the parameters $\gamma^*$ and $\psi^*$ follow from $\tilde{\gamma}$ and $\tilde{\psi}$ upon absorption of $\kappa$, and $c^{**}(\cdot, \cdot)$ is the corresponding normalizing function. Using the two-stage approach (3.6) with (3.37) and (3.39), the marginal model, in analogy with (3.9), equals:

$$f(y \mid \kappa) = \exp\left\{c(y, \phi) + c^{**}(\gamma^*, \psi^*) + c^{**}(\phi^{-1} + \gamma^*, \frac{\phi^{-1} \cdot y + \gamma^* \cdot \psi^*}{\phi^{-1} + \gamma^*})\right\}, \tag{3.41}$$

When strong conjugacy holds for the marginal model $f(y \mid \kappa)$, the marginal joint distribution $f(y)$ and its corresponding moments are easy to compute. However, marginalization is not guaranteed when the strong conjugacy principle does not hold. For example,

Bernoulli models with logit link do not allow for strong conjugacy with the beta and normal random effects and explicit expressions for the marginal moments and joint marginal distribution are unavailable. To handle this issue, generic approximations can be used for the marginal moments (Section 3.4.2), and/or even numerical integration techniques such as adaptive Gaussian quadrature and ordinary Gaussian quadrature (Molenberghs and Verbeke, 2005) to derive such marginal quantities. For example, Ivanova et al (2014) used the latter one to provide accurate parameter estimates for a combined proportional odds-Beta-normal model in an overdispersed hierarchical ordinal setting. A special case is obtained when using the probit link in Bernoulli models, that allows for closed-form marginalization, even though strong conjugacy does not apply. Therefore, it can be said that closed-form marginalization are present when the strong conjugacy principle applies, but not vice versa.

## 3.5 General Estimation Strategies

In this section, a general exploration is given to the frequentist and Bayesian approaches (Cox, 2006; Stegmueller, 2013). These approaches are principal views which help to draw statistical inference, i.e., a procedure with the aim to extract information from collected data by generalizing the observed results beyond the sample data to a population or to the future. The frequentist viewpoint is based only on the observed data from the current experiment, e.g., maximum likelihood (Section 4.2.1) and pairwise likelihood (Section 4.2.2), while the Bayesian perspective also accommodates learning from previous experiments and/or previous evidence. Explaining these approaches are essential for fully understanding the proposed estimation strategies of the CM for the time-to-event setting (Section 4.2). Both methods often lead to the same solution when no external information (other than the data and the model itself) is introduced into the analysis.

### 3.5.1 Frequentist Estimation Approach

The frequentist approach assumes a distribution of a random variable $\mathbf{Y}$, governed by a parameter vector $\boldsymbol{\iota}$. In case of a normal distribution, this parameter vector equals $\boldsymbol{\iota}$ equals $(\boldsymbol{\mu}, D)$ with $\boldsymbol{\mu}$ the mean vector and $D$ the variance-covariance matrix. The aim is to estimate the unknown parameter vector $\boldsymbol{\iota}$ by randomly selecting an appropriate sample $\mathbf{y} = (y_1, \ldots, y_n)$. To achieve this goal, different procedures, e.g., the least squares estimation (LSE) or maximum likelihood estimation (MLE), can be used. The former one has been a popular choice of model fitting in psychology (e.g., Rubin et al, 1999; Lamberts, 2000; Myung, 2003) and is tied to many familiar statistical concepts such as linear regression, sum of squared error (SSE), proportion variance accounted for (i.e. $r^2$), and root mean squared deviation (RMSD). Unlike MLE, introduced by Fisher in the 1920s (Stigler, 2007), LSE requires no or minimal distributional assumptions. It is useful for obtaining a descriptive measure for the purpose of summarizing observed data, but has no basis for testing hypotheses or constructing confidence intervals. The latter one, the MLE, owns many optimal properties in estimation e.g., sufficiency (complete information about the parameter of interest contained in its MLE estimator); consistency (true parameter value that generated the data recovered asymptotically, i.e. for data of sufficiently large

samples); efficiency (lowest-possible variance of parameter estimates achieved asymptotically); and parameterization invariance (same MLE solution obtained independent of the parametrization used), which aren't present in LSE. Further, many of the inference methods in statistics are developed based on MLE. For example, MLE is a prerequisite for the chi-square test, Bayesian methods, inference with missing data, modeling of random effects, and many model selection criteria such as the Akaike information criterion (Akaike, 1973 and 1974) and the Bayesian information criteria (Schwarz, 1978).

In this thesis, the MLE principle is discussed to make inferences on the proposed CM (Section 4.2.1). MLE maximizes the likelihood function $L(\boldsymbol{\iota} \mid y_1, \ldots, y_n) = f(y_1, \ldots, y_n \mid \boldsymbol{\iota})$ where $f(\cdot)$ represents the probability density function of $\mathbf{Y}$. In practice, it is more computational convenient to work with the logarithm of the likelihood function $l(\boldsymbol{\iota} \mid y_1, \ldots, y_n) = \ln[L(\boldsymbol{\iota} \mid y_1, \ldots, y_n]$, due to the parameterization invariance property. However, in the CM framework, maximization of the likelihood (3.36) need to be done. The main problem in maximizing (3.36) is the presence of $N$ integrals over the random effects $\mathbf{b}_i$ and $\boldsymbol{\theta}$. In some special cases, e.g. the linear mixed model for continuous outcomes (Section 3.3.1) where likelihood (3.31) is used, the $N$ integrals can be worked out analytically. However, in general, no analytic expressions are available for the integrals in (14.2) and numerical approximations are claimed (Molenberghs and Verbeke, 2005).

Molenberghs and Verbeke (2005) discussed a number of numerical approximations and subdivided them in those that are based on the approximation of the integrand, those based on an approximation of the data, and those that are based on the approximation of the integral itself, explaining the popularity of Taylor-series expansion based methods, such as penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL), Laplace approximation, and numerical-integration based methods (e.g., Gaussian quadrature and adaptive Gaussian quadrature). An extensive overview of many approximations can be found in Tuerlinckx et al (2004), Pinheiro and Bates (2000), and Skrondal and Rabe-Hesketh (2004). Several of the series expansion methods tend to exhibit bias, an issue taken up in Breslow and Lin (1995), and suggesting the use of alternative methods.

Additionally, closed-form integration, apart from the normal case, is within reach for the Poisson, probit and Weibull cases. Now, some closed forms involve series expansions, and may be either time consuming or cumbersome to implement. This notwithstanding, a variety of alternative approaches are possible, such as partial marginalization (Molenberghs et al, 2010) and pairwise likelihood (Molenberghs and Verbeke, 2005; Renard et al, 2004). These approaches are discussed in the time-to-event setting (Section 4.2.1 and 4.2.2).

### 3.5.2 Bayesian Estimation Approach

Alternatively to the frequentist paradigm, the Bayesian estimation approach can be applied (Berger, 2006; D'Agostini, 2003; Swinburne, 2002). This approach differs from the frequentist methodology in the way that it incorporates previous evidence of the parameter in addition to the observed data. Therefore, the parameter $\boldsymbol{\iota}$ is assumed to be random with some prior distribution $p(\boldsymbol{\iota})$. Here, similar to the frequentist view, a true value of the parameter $\boldsymbol{\iota}$ is assumed, while the associated prior distribution expresses the uncertainty on that true value. Either informative or non-informative prior can be obtained.

An informative prior expresses specific, definite information about parameter $\boldsymbol{\iota}$, whereas a non-informative prior expresses vague information about parameter $\boldsymbol{\iota}$.

In the Bayesian framework, computation of the posterior distribution, denoted by $p(\boldsymbol{\iota} \mid y_1, \ldots, y_n)$, is of main interest. This distribution is obtained by updating the likelihood $L(\boldsymbol{\iota} \mid y_1, \ldots, y_n)$ (i.e., the observed data) with the prior distribution $p(\boldsymbol{\iota})$ (i.e., previous evidence of the parameter $\boldsymbol{\iota}$, either informative or non-informative). Using Bayes' Theorem, the posterior distribution can be calculated as

$$p(\boldsymbol{\iota} \mid y_1, \ldots, y_n) = \frac{L(\boldsymbol{\iota} \mid y_1, \ldots, y_n) \cdot p(\boldsymbol{\iota})}{p(y_1, \ldots, y_n)} = \frac{L(\boldsymbol{\iota} \mid y_1, \ldots, y_n) \cdot p(\boldsymbol{\iota})}{\int L(\boldsymbol{\iota} \mid y_1, \ldots, y_n) \cdot p(\boldsymbol{\iota}) \cdot d\boldsymbol{\iota}}, \tag{3.42}$$

Because the denominator in (3.42), often called the normalizing constant, is depending only on the observed data $y_1, \ldots, y_n$ (assumed to be fixed in the Bayesian methodology), the posterior distribution is proportional to the prior distribution and likelihood, i.e.,

$$p(\boldsymbol{\iota} \mid y_1, \ldots, y_n) \propto L(\boldsymbol{\iota} \mid y_1, \ldots, y_n) \cdot p(\boldsymbol{\iota}), \tag{3.43}$$

However, a major limitation towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution $p(\boldsymbol{\iota} \mid y_1, \ldots, y_n)$ often requires the integration of high-dimensional functions in the normalizing constant. This often presents computational difficulties, but several approaches short of direct integration have been proposed (Smith, 1991; Evans and Swartz, 1995). This thesis focuses on the Markov Chain Monte Carlo (MCMC) methods (Gilks, 2005; Geyer, 1992; Lesaffre and Lawson, 2012), which attempt to simulate direct draws from some complex distribution of interest.

MCMC is a sampling technique where values of $\boldsymbol{\iota}$ are drawn sequentially from an approximate distribution and correcting the samples to better approximate the target posterior distribution, $p(\boldsymbol{\iota} \mid y_1, \ldots, y_n)$. The sequentially draws represent a Markov chain, where the distribution of the sampled value depends on the most recent value drawn, i.e., generating $\boldsymbol{\iota}^{(1)}, \ldots, \boldsymbol{\iota}^{(T)}$ such that $f(\boldsymbol{\iota}^{(t)} \mid \boldsymbol{\iota}^{(t-1)}, \ldots \boldsymbol{\iota}^{(1)}) = f(\boldsymbol{\iota}^{(t)} \mid \boldsymbol{\iota}^{(t-1)})$. Thus, an iterative procedure is preformed where the approximate distribution is improved at each step $(t)$ in the simulation, in the sense of converging to the target posterior distribution. Hence, the distribution of $\boldsymbol{\iota}^{(t)}$ converges to the target posterior distribution as $t \to \infty$. The two most popular MCMC sampling methods are the Gibbs sampler (Geman and Geman, 1984) and the Metropolis-Hastings algorithm (Chib, 1995; Hastings, 1970). A brief discussion about these two methods can be found in Carlo (2004).

Unfortunately, the values are not immediately drawn from the posterior distribution. An initial part, also known as the burn-in part, needs to be discarded and it is absolutely necessary to check the convergence of the sampled sequences. Convergence can be explored by using an informal or a formal check. A trace plot is an informal check of the convergence of the MCMC samples which indicates how quickly the sampling procedure explore the posterior distribution (Gelman et al, 2004). On the other hand, the Gelman-Brooks-Rubin diagnostic is a formal check of the convergence by comparing the between and within-sequence variances (Gelman and Rubin, 1992; Gelman et al, 2004).

# Chapter 4

# A Combined Model for Time-to-Event Data

In this chapter, emphasis is given on the modeling framework (Section 3.4) in an overdispersed, hierarchical time-to-event setting, by build upon the combined model of Molenberghs et al (2010) and Ghebretinsae et al (2011) with the possibility to accommodate for censorship. Censorship, either informative or non-informative, gives rise to an extra complexity. Additionally, different types of censoring (e.g., right-, left-, and interval censoring) are possible. Here, non-informative right censoring is assumed, while informative censoring mechanisms are taking into account in a joint marginalized multilevel model, i.e. JMMM (Section 6.5.2). A similar implementation procedure is possible for left- and interval censoring.

In Section 4.1, a hierarchical formulation is given for the proposed CM. This is supplemented with different estimation approaches for the fixed effects and variance components. While Molenberghs et al (2010) focused on maximum likelihood, using so-called partial marginalization (Section 4.2.1), additionally methods with pairwise likelihood ideas (Molenberghs and Verbeke, 2005) and Bayesian estimation, particularly MCMC (Section 3.5.2), are provided. Often, researchers are also interested in subject-specific inferences by estimating the random effects. Laird and Ware (1982) and Molenberghs and Verbeke (2005) introduced the so-called Empirical Bayes (EB) estimation, and will be discussed in Section 4.3 for the proposed CM framework (Section 4.1).

However, in case that high-dimensional hierarchical structures are present in the data, estimation will become increasingly complex. To resolve this issue, the Alternating Imputation Posterior (AIP), introduced by Clayton and Rashbash (1999) and further studied by Ecochard an Clayton (2002), is proposed (Section 4.4).

## 4.1 The Proposed Combined Model

Given the focus on time-to-event outcomes in an overdispersed, hierarchical data structure, where the hierarchy is made up of repeated events and other types of clustering, simultaneously, it is natural to select the Weibull distribution for (3.32) and the Gamma distribution for $\Upsilon_{ij}$ in Section 3.4.1. This choice is motivated in Section 3.2, i.e., that conjugacy holds. Assuming its mean $\varpi_{ij} \equiv \varpi_j$ and variance $\sigma^2_{ij} \equiv \sigma^2_j$ constant across

clusters (subjects) and re-parameterizing it using the conventional gamma-distribution parameters, gives $\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j)$. With these choices, and adding normal random effects for the 'GLMM part', the proposed model can be formulated by

$$f(\mathbf{y}_i \mid \boldsymbol{\theta}_i, \mathbf{b}_i) = \prod_{j=1}^{n_i} \lambda \cdot \rho \cdot \theta_{ij} \cdot y_{ij}^{\rho-1} \cdot e^{\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i} \cdot e^{-\lambda \cdot y_{ij}^{\rho} \cdot \theta_{ij} \cdot e^{\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i}}, \tag{4.1}$$

$$f(\boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} \frac{1}{\beta_j^{\alpha_j} \cdot \Gamma(\alpha_j)} \cdot \theta_{ij}^{\alpha_j - 1} \cdot e^{-\theta_{ij}/\beta_j}, \tag{4.2}$$

$$f(\mathbf{b}_i) = \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i}. \tag{4.3}$$

i.e., the conditional outcome, conjugate, and normal random effects distribution, respectively. The same terminology is used as before (Section 3.1), where $\lambda$ and $\rho$ are the conventional Weibull shape and scale parameters. It is implicit that the Gamma random effects are independent. The modeling framework (4.1)–(4.3), also known as the so-called Weibull-Gamma-normal (WGN) model (Molenberghs et al, 2010), enjoys the property of strong conjugacy (Section 3.4.3), i.e., conjugacy still applies even after incorporating normal random effects. This is owed by the fact that conjugacy holds for a Gamma distribution in the Weibull case (Section 3.2) and the following property of the Gamma distribution:

$$\begin{aligned} \frac{1}{\kappa} \cdot f(\theta \mid \alpha, \beta) &= \frac{1}{\kappa} \cdot \frac{1}{\beta^{\alpha} \cdot \Gamma(\alpha)} \cdot \theta^{\alpha-1} \cdot e^{-\theta/\beta} \\ &= \frac{1}{(\kappa \cdot \beta)^{\alpha} \cdot \Gamma(\alpha)} \cdot (\kappa \cdot \theta)^{\alpha-1} \cdot e^{-(\kappa \cdot \theta)/(\kappa \cdot \beta)} \\ &= f(\kappa \cdot \theta \mid \alpha, \kappa \cdot \beta), \end{aligned} \tag{4.4}$$

Therefore, closed-form expressions can be derived for the marginal joint distribution, mean, variance, and higher-order moments. These derivations can be found in Appendix B.1, where also a number of related facts are derived. From the WGN model, special cases such as the classical frailty model (i.e., no normal-random effects) and the Weibull-based GLMM (i.e., no Gamma random effects) follow. Additionally, both conjugate random effects $\theta_{ij}$ and normal random effects $\mathbf{b}_i$ are here assumed independent of each other.

Alternatively, for simplicity, the WGN model (4.1)–(4.3) can be re-formulated by

$$Y_{ij} \mid \mathbf{b}_i, \theta_{ij} \sim \text{Weibull}(\rho, k_{ij}), \tag{4.5}$$

$$k_{ij} = \lambda \cdot \theta_{ij} \cdot e^{\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i}, \tag{4.6}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \tag{4.7}$$

$$\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j). \tag{4.8}$$

Both formulations (4.1)–(4.3) and (4.5)–(4.8) of the WGN model can be used. Here, the former one will be used to cover up the estimation strategies from Section 4.2 and 4.3, while the latter one is obtained to extend the WGN model to more complex clustering formats (Section 4.4).

## 4.2 Estimation of fixed effects

The attendance of analytically closed-form expressions for the WGN model (Appendix B.1) makes it possible to perform parameter estimation through maximum likelihood. However, while the marginal probabilities could be directly specified for estimation, the existence of infinite series, as can be seen from expressions $(B.8)$–$(B.9)$ in Appendix B.1, may make the approach intractable. Due to its ease of analytical integration over the conjugate random effects, and the availability of software that can numerically integrate over normal random effects, e.g. the SAS procedure NLMIXED (Griswold and Zeger, 2004), a convenient estimation route can be provided by combining analytical integration and numerical integration. Molenberghs et al (2007, 2010) used this ease to introduce the so-called partial marginalization approach, i.e., analytically integrating the hierarchical density (3.32) over the conjugate random effects, and leaving the normal random effects untouched for numerical integration. Therefore, only the expressions for the joint distribution marginal over the conjugate but conditional on the normal random effects need to have a closed-form for estimation.

Alternative to this full likelihood method, focus will be laid on pairwise likelihood (Renard et al, 2004) and Bayesian estimation for the WGN model. Both are extensively discussed in Molenberghs et al (2012) and Efendi and Molenberghs (2013), while Ghebretinsae et al (2013) placed emphasis on the latter one.

### 4.2.1 Maximum Likelihood

Maximum likelihood estimation with partial marginalization (Molenberghs et al, 2007 and 2010) in the time-to-event case, when using the WGN model, analytically integrates out the hierarchical density (4.1) over the Gamma random effects, leaving the normal random effects to numerical integration. However, in the survival case, it is often likely that censoring occurs. Avoiding this concept in the estimation produce may lead to bias results, resulting in wrong conclusions. Focusing on the right-censored data, for each $j$, the conditional distribution is integrated over the time interval $[C_{ij}, +\infty[$. Here, $C_{ij}$ denotes the $j$th censored time for cluster (subject) $i$. Additionally, the occurrence of censoring on cluster (subject) $i$ is declared by the censoring indicator $\delta_i$, which equals 1 if $y_{ij} \leq C_{ij}$ and 0 if $y_{ij} > C_{ij}$. The corresponding marginal-conditional density in the WGN model equals:

$$f(y_{ij} \mid \mathbf{b}_i) = \frac{(\lambda \cdot e^{\eta_{ij}} \cdot \rho \cdot y_{ij}^{\rho-1} \cdot \alpha_j \cdot \beta_j)^{\delta_i}}{(1 + \lambda \cdot y_{ij}^{\rho} \cdot e^{\eta_{ij}} \cdot \beta_j)^{\alpha_j + \delta_i}}, \tag{4.9}$$

with $\eta_{ij} = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i$ and the same terminology of parameters is used as before (Section 4.1). The marginal likelihood for $\boldsymbol{\xi}, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}$, denoted by $L(\boldsymbol{\xi}, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta})$, is derived as

$$L(\boldsymbol{\xi}, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{N} \left[ \int_b \left( \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i \right) \right]. \tag{4.10}$$

To integrate out the normal random effects in the marginal likelihood (4.10), numerical-integration based methods, such as adaptive Gaussian quadrature (Molenberghs and Ver-

beke, 2005), are used. Parameter estimates are than obtained by maximizing the obtained (approximated) marginal likelihood.

The concept of partial integration always applies whenever strong conjugacy holds. Indeed, an expression of the form (3.41) corresponds to integrating over the conjugate random effect $\theta$, while leaving the normally distributed random effect embedded in the predictor, $\kappa$ in this notation. Recall that, while expressions of the type (3.41) appear to be for the univariate case, they extend without problem to the longitudinal setting as well.

### 4.2.2 Pairwise Likelihood

As an alternative to full likelihood in the frequentist approach, so-called pairwise-likelihood estimation (Renard, Molenberghs, and Geys, 2004) can be used. Since pairwise likelihood is a special case of pseudo-likelihood (Aerts et al, 2002; Arnold and Strauss, 1991; Molenberghs et al, 2012; Molenberghs and Verbeke, 2005), attention will first be given on the latter one and specified into the former one.

The principal idea behind pseudo-likelihood is to replace a numerically challenging joint marginal distribution by a simpler function assembled from suitable factors. This strategy is useful when the computational burden of full likelihood becomes burdensome and/or when robustness against misspecification of higher-order moments is desirable. This is especially the case when the joint marginal distribution is available but cumbersome to manipulate and evaluate, such as the joint marginal distribution ($B.8$) of the WGN model. It can also stabilize computations and make the iterative process less dependent on starting values, even though it may not always reduce computation time. Essentially, the joint distribution is replaced with a product of factors of marginal and/or conditional distributions of lower dimensions. Because such a product does not necessarily re-compose the original joint distribution, sandwich-estimator ideas are then used to provide not only valid point estimates, but also precision estimates and inferences derived therefrom (Molenberghs and Verbeke, 2005). Next, a general (mathematically) discussion is given of this approach.

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

$$pl = \sum_{i=1}^{N} \sum_{s \in S} \delta_s \cdot \ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]. \tag{4.11}$$

Adequate regularity conditions have to be invoked to ensure that (4.11) can be maximized by solving the pseudo-likelihood (score) equations, the latter obtained by differentiating the logarithmic pseudo-likelihood and by equating its derivative to zero. These regularity conditions are given in Appendix B.2. In particular, when the components in (4.11) result from a combination of marginal and conditional distributions of the original distribution,

then a valid pseudo-likelihood function results. Additionally, the classical log-likelihood function is found by setting $\delta_s = 1$ if $s$ is the vector consisting solely of ones, and 0 otherwise. More details can be found in Varin (2008), Lindsay (1988), and Joe and Lee (2008). Note that Joe and Lee (2008) use weighting for reasons of efficiency in pairwise likelihood, similar in spirit to Geys, Molenberghs, and Lipsitz (1998), but differently from its use here, which focuses on bias correction when data are incomplete. Another important reference is Cox and Reid (2004).

Let $\boldsymbol{\xi}_0$ be the true parameter vector. Under the suitable regularity conditions (**B1**)–(**B7**), Arnold and Strauss (1991) showed that maximizing the function (4.11) produces a consistent and asymptotically normal estimator $\tilde{\boldsymbol{\xi}}_N$ so that $\sqrt{N} \cdot (\tilde{\boldsymbol{\xi}}_N - \boldsymbol{\xi}_0)$ converges in distribution to $N_p(\mathbf{0}, I_0(\boldsymbol{\xi}_0)^{-1} \cdot I_1(\boldsymbol{\xi}_0) \cdot I_0(\boldsymbol{\xi}_0)^{-1})$, where $I_0$ and $I_1$ are defined in $(B.20)$ and $(B.21)$ respectively. More details can be found in Appendix B.2.

As stated earlier, models for non-Gaussian data can become prohibitive when subjected to full maximum likelihood inference, especially with a lot of within-cluster replication. le Cessie and van Houwelingen (1991) and Geys, Molenberghs, and Lipsitz (1998) replace the true contribution of a vector of correlated binary data to the full likelihood, written as $f(y_{i1}, \ldots, y_{in_i})$, by the product of all pairwise contributions $f(y_{ij}, y_{ik})$ $(1 \leq j < k \leq n_i)$, to obtain a pseudo-likelihood function. Also the term *composite likelihood* is encountered in this context. Renard, Molenberghs, and Geys (2004) refer to this particular instance of pseudo-likelihood as *pairwise likelihood*. Grouping the outcomes for subject $i$ into a vector $\mathbf{Y}_i$, the contribution of the $i$th cluster to the log pseudo-likelihood then specializes to

$$pl_i = \sum_{j<k} \ln \left[ f(y_{ij}, y_{ik}) \right]. \tag{4.12}$$

if it contains more than one observation. Otherwise $pl_i = f(y_{i1})$. Extension to three-way and higher-order pseudo-likelihood is straightforward. All of these are special cases of (4.11). Renard, Molenberghs and Geys (2004) used this estimation procedure for estimating multilevel probit models with random effects, while Molenberghs et al (2012) and Efendi, Molenberghs and Iddi (2013) compared this approach with the conventional maximum full likelihood principle (Section 4.2.1) to estimate recurrent asthma attacks in children in the recurrent asthma data (Duchateau and Janssen, 2007), where censoring is present, with the proposed WGN model from Section 4.1. Additionally, the former one also compared these procedures with the WGN model in the comet assay data (Molenberghs and Verbeke, 2005), where no censoring is present.

### 4.2.3 Bayesian estimation

Alternatively to the previous frequentist approaches (Section 4.2.1 and 4.2.2), a fully Bayesian route can be followed. This does not only have computational advantages, it also allows to take relevant information from preceding studies into account by updating the prior belief of the parameters' distribution. In the last few years, the Bayesian approach has become popular among researchers, especially for fitting the CM. For example, Aregay, Shkedy, and Molenberghs (2013) studied the CM from a Bayesian perspective for

analyzing the epilepsy data set (Thall and Vail, 1990), while Ghebretinsae et al (2013) extended the WGN model to the Weibull-Gamma-normal-normal (WGNN) model, i.e., adding an extra normal random effect to the WGN model, to accommodate the full hierarchical structure of the comet assay data and analyzed it in a Bayesian approach.

In the WGN case, using the same terminology as before, formula (3.43) becomes

$$f(\boldsymbol{\vartheta}, D, \boldsymbol{\varpi}, \Sigma \mid \mathbf{y}) \propto \prod_{i=1}^{N} \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i) \cdot f(\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i), \tag{4.13}$$

with prior densities $f(\boldsymbol{\vartheta}, \mathbf{b}_i, \boldsymbol{\theta}_i)$. Inferences are typically made by taking random draws from the posterior density $f(\boldsymbol{\vartheta}, D, \boldsymbol{\varpi}, \Sigma \mid \mathbf{y})$ using MCMC. A further discussion on this topic can be found in Section 3.5.2. Unlike the popularity of the SAS procedure NLMIXED in the frequentist approach, other statistical programs such as Winbugs, R2Winbugs, and R2jags are commonly used in the Bayesian framework. These packages allow for MCMC based integration over multiple random effects (Browne and Draper, 2002).

## 4.3 Estimation of the random effects

Although in practice one is usually primarily interested in estimating the parameters of the marginal distribution ($B.8$), i.e., $\boldsymbol{\xi}$, $D$, $\lambda$, $\rho$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, estimations of the random effects $\theta_{ij}$ and $\mathbf{b}_i$ are often useful as well, e.g., for detecting special profiles (i.e., outlying individuals) or groups of individuals evolving differently in time (Chapter 9). Additionally, estimates for the random effects are substantially for predicting subject-specific evolutions. Because the parameters $\theta_{ij}$ and $\mathbf{b}_i$ are assumed to be random, it is most natural to estimate them using Bayesian techniques (Box and Tiao, 1992; Gelman et al, 1995). Laird and Ware (1984) and Molenberghs and Verbeke (2005) introduced so-called Empirical Bayes (EB) estimation, which combines a Bayesian approach with maximum likelihood. Even though the latter one focuses on EB estimations in the LMM and GLMM settings, this thesis addresses an EB estimation strategy for the CM framework (Iddi, Molenberghs, Aregay, and Kalemab, 2014), particular for the WGN model. Iddi et al (2014) used this EB estimation approach to evaluate individual profiles in the epilepsy data set and recurrent asthma data, where the WGN model is used for the latter one.

### 4.3.1 Empirical Bayes Estimation of Normal Random Effect

While Molenberghs and Verbeke (2005) explored the EB estimation for the LMM and GLMM setting, Iddi et al (2014) extended this in the CM framework, where both random effects $\theta_{ij}$ and $\mathbf{b}_i$ need to be estimate. The procedure of EB estimation uses maximum likelihood estimates to replace the unknown parameters in the posterior distribution of the random effect. In case of the normal random effect $\mathbf{b}_i$, the density for the posterior distribution $\mathbf{b}_i$ is given by

$$f(\mathbf{b}_i \mid \mathbf{y}_i) = \frac{f(\mathbf{y}_i \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i)}{\int_b f(\mathbf{y}_i \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}, \tag{4.14}$$

where the parameters are suppressed from notation. The distributions involved in (4.14) are marginalized over the random effects $\theta_{ij}$. These obtain closed forms in the WGN setting, due to the strong conjugacy principle (3.41). By using the same terminology as before, the following posterior densities $f(\mathbf{b}_i \mid \mathbf{y}_i)$ are given for the WGN model:

$$f(\mathbf{b}_i \mid \mathbf{y}_i) \propto \prod_{j=1}^{n_i} \frac{(\lambda \cdot e^{\eta_{ij}} \cdot \rho \cdot y_{ij}^{\rho-1} \cdot \alpha_j \cdot \beta_j)^{\delta_i}}{(1+\lambda \cdot y_{ij}^{\rho} \cdot e^{\eta_{ij}} \cdot \beta_j)^{\alpha_j+\delta_i}} \cdot \frac{1}{(2\cdot\pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot e^{\frac{1}{2}\cdot \mathbf{b}_i' \cdot D^{-1} \cdot \mathbf{b}_i}, \tag{4.15}$$

The estimates of the random effects based on the posterior distribution are obtained from:

$$\widehat{\mathbf{b}}_i = \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) = \int_b \mathbf{b}_i \cdot f(\mathbf{b}_i \mid \mathbf{y}_i) \cdot d\mathbf{b}_i. \tag{4.16}$$

Alternatively, the random effects $\mathbf{b}_i$ are sometimes treated as unknown parameters to be estimated by the following maximization problem:

$$\widehat{\mathbf{b}}_i = \mathrm{argmax}_{\mathbf{b}_i} f(\mathbf{b}_i \mid \mathbf{y}_i).$$

Although (4.16) can easily be programmed, the estimates and their standard errors are also given as a bonus when using a tool such as the SAS procedure NLMIXED, as long as the correct conditional densities involving only the normal random effect are specified.

### 4.3.2 Empirical Bayes Estimation of Conjugate Random Effects

A similar approach can be followed for $\theta_{ij}$. The posterior density for $\theta_{ij}$ equals

$$f(\theta_{ij} \mid y_{ij}) = \frac{f(y_{ij} \mid \theta_{ij}) \cdot f(\theta_{ij})}{\int_\theta f(y_{ij} \mid \theta_{ij}) \cdot f(\theta_{ij}) \cdot d\theta_{ij}}, \tag{4.17}$$

with $f(y_{ij} \mid \theta_{ij}) = \int_b f(\mathbf{y}_i \mid \theta_{ij}, \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i$ and the parameters are suppressed from notation. From Appendix B, the conditionally densities $f(\mathbf{y}_i \mid \boldsymbol{\theta}_i)$ of the WGN model (i.e., formula $(B.3)$) involves sums of infinite series that are cumbersome to compute. Therefore, it is handier to carry out this integration numerically (e.g., adaptive Gaussian quadrature). The challenge is the intensive computation involved, especially when the dimension of $\mathbf{b}_i$ and $\theta_{ij}$ combined is high. To handle this difficulty, Iddi et al (2014) proposed a two-stage approach, where the random effects $\mathbf{b}_i$ are estimated (according to Section 4.3.1) and replaced them in (4.17) in the first stage. The parameters $\boldsymbol{\xi}, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}$ are also estimated and treated as known parameters in (4.17). In the second stage, the estimator $\widehat{\theta}_{ij}$ of $\theta_{ij}$ is obtained from:

$$\widehat{\theta}_{ij} = \mathrm{E}(\theta_{ij} \mid y_{ij}) = \int_\theta \theta_{ij} \cdot f(\theta_{ij} \mid y_{ij}) \cdot d\theta_{ij}. \tag{4.18}$$

Iddi et al (2014) motivated this approach by developing an appropriate SAS macro for handling this computation (Appendix B.3). Estimation of the standard errors are done in a similar way, using the posterior distribution. From

$$\mathrm{Var}(\widehat{\theta}_{ij}) = \int_\theta (\theta_{ij} - \widehat{\theta}_{ij}) \cdot (\theta_{ij} - \widehat{\theta}_{ij})' \cdot f(\theta_{ij} \mid y_{ij}) \cdot d\theta_{ij}, \tag{4.19}$$

the standard errors are obtained by taking the square root of $\mathrm{Var}(\widehat{\theta}_{ij})$.

## 4.4 Extension to more complex clustering formats

Until now, only one hierarchical level is considered in the proposed WGN model (4.5)–(4.8). This assumption is often inadequate to accommodate the full hierarchical structure. For example, choosing only one hierarchical level falls short of analyzing the comet assay data. Therefore, extensions of the CM model are needed to cover up this higher hierarchical structures. A straightforward choice is to propose more than one normal random effect in the model, where the number of normal random effects equals the amount of hierarchical levels.

In case of two hierarchical levels, the WGN model (4.5)–(4.8) becomes:

$$Y_{i_1i_2i_3} \mid \mathbf{b}_{i_1}, \mathbf{b}_{i_1i_2}, \theta_{i_1i_2i_3} \sim \text{Weibull}(\rho, k_{i_1i_2i_3}), \tag{4.20}$$

$$k_{i_1i_2i_3} = \lambda \cdot \theta_{i_1i_2i_3} \cdot e^{\mathbf{x}'_{i_1i_2i_3}\cdot\boldsymbol{\xi}+\mathbf{z}'_{i_1i_2}\cdot\mathbf{b}_{i_1}+\mathbf{z}'_{i_1i_2i_3}\cdot\mathbf{b}_{i_1i_2}}, \tag{4.21}$$

$$\mathbf{b}_{i_1} \sim N(\mathbf{0}, D_1), \tag{4.22}$$

$$\mathbf{b}_{i_1i_2} \sim N(\mathbf{0}, D_2), \tag{4.23}$$

$$\theta_{i_1i_2i_3} \sim \text{Gamma}(\alpha_{i_3}, \beta_{i_3}), \tag{4.24}$$

with observation $i_3$ in sub-cluster $i_2$, located in the (top-)cluster $i_1$ ($i_1 = 1, \ldots, N; i_2 = 1, \ldots, n_{i_1}; i_3 = 1, \ldots, n_{i_1i_2}$). Efendi and Molenberghs (2013) named this two-cluster-level model (4.20)–(4.24) the Weibull-Gamma-normal-normal (WGNN) model, which is one of the four hierarchical overdispersed Weibull models that were analyzed in Ghebretinsae et al (2011) for the comet assay data.

Generally, extending the WGN model to a multilevel CM with $m$ cluster levels can be done in a similar way, leading to:

$$Y_{i_1i_2\ldots i_m} \mid \mathbf{b}_{i_1}, \mathbf{b}_{i_1i_2}, \ldots, \mathbf{b}_{i_1i_2\ldots i_{m-1}}, \theta_{i_1i_2\ldots i_m} \sim \text{Weibull}(\rho, k_{i_1i_2\ldots i_m}), \tag{4.25}$$

$$k_{i_1i_2\ldots i_m} = \lambda \cdot \theta_{i_1i_2\ldots i_m} \cdot e^{\mathbf{x}'_{i_1i_2\ldots i_m}\cdot\boldsymbol{\xi}+\sum_{l=1}^{m-1}\mathbf{z}'_{i_1\ldots i_l}\cdot\mathbf{b}_{i_1\ldots i_l}}, \tag{4.26}$$

where $\theta_{i_1i_2\ldots i_m}$ is Gamma distributed with parameters $\alpha_{i_m}$ and $\beta_{i_m}$ and all $\mathbf{b}_{i_1\ldots i_l}$ are normal distributed with mean $\mathbf{0}$ and variance $D_l$. Unfortunately, estimation will become increasingly complex with growing $m$. To resolve this issue, Efendi and Molenberghs (2013) proposed to make use of the Alternating Imputation Posterior (AIP) algorithm, introduced by Clayton, and Rasbash (1999) and further studied by Ecochard, and Clayton (2002) and Cho, and Rabe-Hesketh (2011). A general overview of this algorithm can be found in Appendix B.

Applying the AIP algorithm, the model above can be partitioned such that, for $l = 1, \ldots, m$, the $j$th order random-effects structure is fitted by considering

$$o^{(j)} = e^{\sum_{l\neq j}\mathbf{z}'_{i_1\ldots i_l}\cdot\mathbf{b}_{i_1\ldots i_l}}, \tag{4.27}$$

as a known offset, leading formula (4.26) to the $j$th nested sub-model

$$\begin{aligned} k_{i_1i_2\ldots i_m} &= \lambda \cdot \theta_{i_1i_2\ldots i_m} \cdot o^{(j)} \cdot e^{\mathbf{x}'_{i_1i_2\ldots i_m}\cdot\boldsymbol{\xi}+\mathbf{z}'_{i_1\ldots i_j}\cdot\mathbf{b}_{i_1\ldots i_j}} \\ &\propto \lambda \cdot \theta_{i_1i_2\ldots i_m} \cdot e^{\mathbf{x}'_{i_1i_2\ldots i_m}\cdot\boldsymbol{\xi}+\mathbf{z}'_{i_1\ldots i_j}\cdot\mathbf{b}_{i_1\ldots i_j}}. \end{aligned} \tag{4.28}$$

At iteration step $(t)$, the following steps are carried out, for each $l = 1, \ldots, m$:

1. Fit the $j$th nested sub-model (4.28) using the offset $o^{(j)}$ calculated from current values of the random effects;

2. Sample the model parameters $\boldsymbol{\xi}^{(t)}$ from an approximation to the joint posterior distribution of $\boldsymbol{\xi}$, given $\mathbf{y}$, $\mathbf{x}$ and $o^{(j)}$;

3. Sample $\mathbf{b}^{(t)}_{i_1 \ldots i_j}$ from the posterior distribution of $\mathbf{b}_{i_1 \ldots i_j}$, given $\mathbf{y}$, $\mathbf{x}$ and $o^{(j)}$, which is normally distributed.

The AIP has been implemented in both a sequential as well as a parallel fashion (Ecochard and Clayton, 2002). While the overall parameter ($\boldsymbol{\xi}$, $\mathbf{b}$) is estimated by ($B$.22), the variance is the mean of the variances plus the variance of the estimates, using the law of iterated expectations. By using the principles laid out in Ecochard and Clayton (2002), a Gaussian approximation is obtained using Rao-Blackwellization (Gelfand and Smith, 1990), leading to formula ($B$.23) for the variance.

To end this discussion, it can be said that, within each AIP, a combined model with one normally distributed random effect as well as a conjugate random effect is fit. Estimation of the fixed effects can be done through the three distinct strategies of full likelihood (Section 4.2.1), pairwise likelihood (Section 4.2.2), or Bayesian nature (Section 4.2.3). Furthermore, Efendi and Molenberghs (2013) mentioned that, unlike with maximum and pairwise likelihood, Bayesian estimation does not combine with AIP. Rather, direct implementation is possible using standard statistical programs such as Winbugs, R2Winbugs, and R2jags. These packages allow for MCMC based integration over multiple random effects (Browne and Draper, 2002).

# Chapter 5

# A Marginalized Combined Model for Time-to-Event Data

Even though the parameter estimates from the hierarchical LMM (3.22)–(3.23) have a marginal interpretation, i.e., $\mathrm{E}(\mathbf{Y}_i) = \mathrm{E}[\mathrm{E}(\mathbf{Y}_i \mid \mathbf{b}_i)] = \mathbf{X}_i \cdot \boldsymbol{\xi}$, this is generally not the case for the parameters in the GLMM (3.29)–(3.30) for non-Gaussian outcomes (e.g. binary, time-to-event, etc.), where a subject-specific interpretation is retrieved and not a population-averaged one. For example, using the logit link to relate covariates and random effects to the expectation of a binary outcome $Y_{ij}$,

$$\mathrm{E}(Y_{ij}) = \mathrm{E}[\mathrm{E}(Y_{ij} \mid \mathbf{b}_i)] = E\left[\frac{e^{\mathbf{x}'_{ij}\cdot\boldsymbol{\xi}+\mathbf{z}'_{ij}\cdot\mathbf{b}_i}}{1+e^{\mathbf{x}'_{ij}\cdot\boldsymbol{\xi}+\mathbf{z}'_{ij}\cdot\mathbf{b}_i}}\right] \neq \frac{e^{\mathbf{x}'_{ij}\cdot\boldsymbol{\xi}}}{1+e^{\mathbf{x}'_{ij}\cdot\boldsymbol{\xi}}}. \tag{5.1}$$

As an alternative route, a direct marginal specification can be used to obtain population-averaged interpretations. Liang and Zeger (1986) proposed so-called generalized estimation equations (GEE), which extends GLM by allowing for correlation within cluster (subject) through a so-called working correlation. Even when this working correlation is misspecified, parameter estimators are still consistent and asymptotically normal. However, Molenberghs and Verbeke (2005) and Diggle et al (2002) pointed out that GEE poses challenges when incomplete data is present, i.e., GEE is only valid under missing completely at random (MCAR) and not when data are missing at random (MAR). Many extensions of GEE were suggested in the past. For example, Fitzmaurice et al (2009) extended the GEE framework, based on inverse probability weighting, which allow for MAR. Additionally, this methodology lacks a likelihood basis, which rules out certain inferential routes. A more comprehensive discussion about GEE can be found in Molenberghs and Verbeke (2005), along with the extensions (e.g., Prentice's GEE Method, second-order GEE, etc.).

Choosing a hierarchical or marginal specification depends on the research question. Sometimes, even both are needed. For example, in a clinical trial, researchers are often interested in the marginal effect of a drug, while, at the same time, subject-specific effects of the drug are needed. Therefore, it is useful to provide an elegant framework that simultaneously allows for a marginal as well as a subject-specific interpretation by using both marginal and conditional models at the same time.

In this chapter, emphasis is placed on the so-called marginalized multilevel model (MMM), introduced by Heagerty (1999) and Heagerty and Zeger (2000). This methodology brings

together the strength of both GEE and GLMM, in such a way that the marginal mean, rather than the mean conditional on random effects, is regressed on covariates. The marginal regression parameters are adopted while still permitting individual-level predictions. However, these authors did not provide any connection between the formulated MMM and conditional model. To resolve this issue, Griswold and Zeger (2004) reformulated the MMM to make connections between marginal and conditional models transparent, and additionally construct marginalized models in terms of their conditional model counterparts (e.g., logistic-logistic MMM for the logistic-normal GLMM and probit-probit-normal MMM for the probit-normal GLMM). Unfortunately, they limited this methodology to GLMM, without capturing overdispersion. Iddi and Molenberghs (2012) extended this for the CM framework (Section 3.4) in case of binary and count outcomes.

In Section 5.1, focus will be given on the MMM framework (Heagerty, 1999) for censored, repeated time-to-event outcomes with overdispersion (Efendi, Molenberghs, and Iddi, 2013). A general formulation of this MMM is given in terms of the proposed hierarchical WGN model (4.1)–(4.3) from Section 4.1. Similar estimation strategies (Sections 4.2 and 4.3) can be used. In this context, full likelihood estimation with partial marginalization is discussed (Section 5.2), keeping in mind that pairwise likelihood and Bayesian estimation are possible too.

## 5.1 The Marginalized Combined Model

Even though the CM framework of Section 3.4 accommodates both overdispersion and hierarchical data structures, subject-specific interpretations are retrieved from the fixed-effects vector $\boldsymbol{\xi}$ in (4.1). Knowing the elegant closed-form solution for the marginal mean function of the WGN model (4.1)–(4.3), i.e., formula ($B$.14), Molenberghs et al (2010) showed that, consistent with Zeger, Liang, and Albert (1988), the marginal regression function does not alter, except for the marginal intercept, which depends on the conditional intercept, the scale parameter $\lambda$, and the overdispersion parameters $\alpha_j$ and $\beta_j$. Making inferences on the intercepts, or a combination of covariate effects and intercepts, will become inconvenient. To avoid this, Efendi, Molenberghs and Iddi (2013) proposed a so-called marginalized multilevel model (MMM), in the tradition of Heagerty (1999), Heagerty and Zeger (2000), and Griswold and Zeger (2004), by specifying the regression function marginally and still entering the normal random effects $\mathbf{b}_i$ into the conditional mean function $\mu^c_{ij}$ directly.

Adopting the same terminology as before, this model, also known as the combined overdispersed and marginalized multilevel model (COMMM), takes the form:

$$\mu^m_{ij} = h(\mathbf{x'}_{ij}\boldsymbol{\xi}^m), \tag{5.2}$$

$$\mu^c_{ij} = \theta_{ij} \cdot \kappa_{ij} = \theta_{ij} \cdot h(\Delta_{ij} + \mathbf{z}^{'}_{ij} \cdot \mathbf{b}_i), \tag{5.3}$$

where the fixed-effects parameter vector $\boldsymbol{\xi}^m$ is superscripted to denote its directly marginal interpretation, unlike the fixed-effects parameter vector $\boldsymbol{\xi}$ from Chapter 4, and $\Delta_{ij}$ represents a so-called connector function between (5.2) and (5.3). The predictor on the right hand side of (5.3) replaces the conventional predictor in (4.1).

Due to the dual mean specification (5.1)–(5.2) of COMMM, a defining (closed-form) expression for the connector function $\Delta_{ij}$ (Griswold and Zeger, 2004) follows for the WGN model:

$$\begin{aligned} h(\mathbf{x}'_{ij}\boldsymbol{\xi}^m) = \mu_{ij}^m &= \int_{\theta,b} \theta_{ij} \cdot h(\Delta_{ij} + \mathbf{z}'_{ij} \cdot \mathbf{b}_i) \cdot f(\theta_{ij} \mid \alpha_j, \beta_j) \cdot f(\mathbf{b}_i \mid D) \cdot d\theta_{ij} \cdot d\mathbf{b}_i \\ &= \mathrm{E}(\theta_{ij}) \cdot \int_b h(\Delta_{ij} + \mathbf{z}'_{ij} \cdot \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i \qquad (5.4) \\ &\Downarrow \text{ (WGN model)} \\ \Delta_{ij} &= -\log(\alpha_j \cdot \beta_j) + \mathbf{x}'_{ij}\boldsymbol{\xi}^m - \frac{\mathbf{z}'_{ij} \cdot D \cdot \mathbf{z}_{ij}}{2}. \qquad (5.5) \end{aligned}$$

When no gamma random effects are used in the WGN model, i.e. the Weibull-based GLMM, the first term on the right hand side of (5.5) simply drops. While Efendi, Molenberghs and Iddi (2013) explored this model for the WGN case, Iddi and Molenberghs (2012) focused on binary and count data by proposing appropriate COMMMs with, additional for the latter one, a zero-inflated, overdispersed, and marginalized multilevel model (ZICOMMM).

## 5.2 Estimation and Inference

Similar estimation strategies (Sections 4.2 and 4.3) can be applied for the COMMM, where the predictor on the right hand side of (5.3) replaces the conventional predictor in (4.1). In this section, emphasis is placed on the full likelihood principle for estimating the fixed effects. Other approaches such as Bayesian and pairwise likelihood estimation can be applied too. Additionally, EB estimation of the normal random effects $\mathbf{b}_i$ and gamma random effects $\theta_{ij}$ can be done in a similar way, see Sections 4.3.1 and 4.3.2, respectively.

Full likelihood estimation is provided with reference to Section 4.2.1, where the idea of partial marginalization (Molenberghs et al, 2010) is adopted and right-censoring is taken into account by declaring the occurrence of right-censoring on cluster (subject) $i$ by the censoring indicator $\delta_i$ (either 1 or 0, if $y_{ij} \leq C_{ij}$ or $y_{ij} > C_{ij}$ respectively). The corresponding marginal-conditional density in the COMMM for the Weibull-Gamma-normal case is:

$$f(y_{ij} \mid \mathbf{b}_i) = \frac{(\lambda \cdot e^{\eta_{ij}} \cdot \rho \cdot y_{ij}^{\rho-1} \cdot \alpha_j \cdot \beta_j)^{\delta_i}}{(1 + \lambda \cdot y_{ij}^{\rho} \cdot e^{\eta_{ij}} \cdot \beta_j)^{\alpha_j + \delta_i}}, \qquad (5.6)$$

with $\eta_{ij} = \Delta_{ij} + \mathbf{z}'_{ij} \cdot \mathbf{b}_i = -\log(\alpha_j \cdot \beta_j) + \mathbf{x}'_{ij}\boldsymbol{\xi}^m + \mathbf{z}'_{ij} \cdot \mathbf{b}_i - \frac{\mathbf{z}'_{ij} \cdot D \cdot \mathbf{z}_{ij}}{2}$. The marginal likelihood $L(\boldsymbol{\xi}^m, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is derived as

$$L(\boldsymbol{\xi}^m, D, \lambda, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{N} \left[ \int_b \left( \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i \right) \right]. \qquad (5.7)$$

Implementing this methodology in the SAS procedure NLMIXED can easily be done (Appendix C.1).

# Chapter 6

# Flexible Joint Multilevel Modeling Framework for Repeated, Overdispersed Time-to-Event Data

Over the years, massive attention has been devoted to joint modeling, due to the desire of getting more insight into the data with a single statistical model. Researchers often collect several kinds of outcomes simultaneously in their studies, commonly of a mixed nature. While joint modeling frameworks capture the association between outcomes, separately analysis on these outcomes fails to do it. This chapter focuses on a flexible joint multilevel modeling framework involving an overdispersed, repeated time-to-event setting, by adopting the modeling frameworks from Chapters 4 and 5 for at least one time-to-event outcome and simultaneously capturing the association between outcomes in the correlated normal random effects.

Depending on the research question(s), either subject-specific or population-averaged interpretations are desired. While Section 6.2 places emphasis on subject-specific interpretations, population-averaged interpretations are handled in Section 6.4. Both joint modeling approaches are supported with cases in commonly used real-life situations (Section 6.3 and 6.5, respectively). Whereas the first two cases are related to previously conducted research, i.e., Njagi, Molenberghs et al (2013) and Njagi, Rizopoulos et al (2013), and the third case extends the hierarchical joint modeling strategy of Ghebretinsae et al (2012) to a marginal one, a new marginalized approach (according to the joint marginalized multilevel methodology of Section 6.4) is proposed for capturing informative censoring in an overdispersed, repeated time-to-event setting (Section 6.5.2).

## 6.1 Introduction

In the previously discussed chapters (Chapter 4 and 5), focus has been given to modeling overdispersed, repeated time-to-event outcomes separately by using the so-called WGN model (4.1)–(4.3) for subject-specific interpretations and corresponding COMMM with connector function (5.5) for population-averaged interpretations. However, these types of single modeling strategies are limited in such a way that no answers can be provided to questions relating several or all outcomes simultaneously. For example, in toxicity studies, no single standard endpoint exists 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 these endpoints jointly not only appeals to draw overall inferences using all responses, it also captures the association among the endpoints. In a comet assay, moreover, tail length and tail intensity are commonly used endpoints to assess the DNA damage of a cell as a result of an exposure. Even though univariate analyses are conducted to asses the treatment effect on each endpoint separately, researchers often prefer to reach a conclusion on the overall effect using all outcomes simultaneously, and sometimes even want to know the association between outcomes as well.

While an entire range of combinations is possible in the literature about joint modeling (e.g., continuous/binary, count/survival, etc.), this thesis focuses on four specific cases, each centering on at least one of overdispersed, repeated time-to-event outcome, i.e., two hierarchical joint multilevel modeling frameworks for (1) longitudinal continuous and repeated time-to-event outcomes and (2) repeated binary and repeated time-to-event outcomes, and two marginalized joint multilevel models to accommodate (3) bivariate, repeated time-to-event outcomes and (4) capturing informative censoring mechanisms in an overdispersed, repeated time-to-event setting. While non-informative censoring is assumed in the first three cases, an informative censoring scheme for the time-to-event outcome is proposed in the last case. Furthermore, different joint modeling approaches exist (e.g., Henderson et al, 2000; Tseng et al, 2005; Wintrebert et al, 2005; Fitzmaurice et al, 2009). In this chapter, a flexible multivariate random effects approach, i.e., the so-called extended shared parameter model (ESPM), is chosen for all four cases. In this modeling technique, multivariate random effects are used in which the two outcomes are associated via separate but correlated random effects. The same approach was adopted by Ghebretinsae et al (2012) for modeling two hierarchical, overdispersed time-to-event outcomes jointly in a hierarchical setting. Additionally, Iddi and Molenberghs (2012) used this methodology to model two longitudinal outcomes jointly in a marginalized, multilevel way.

Fieuws and Verbeke (2004) explored this paradigm critically by studying the relation between two hearing thresholds. Here, a discrepancy was found between the data and the relations implied by the joint random-effects model, for the association of the evolutions as well as for the evolution of the association, indicating that it can be misleading to over-interpret the results on the relationship between outcomes implied by the model. Nevertheless, several advantages arise of using random-effects models for joint modeling purposes: (1) the different outcomes do not necessarily need to be of the same type, i.e., either continuous or discrete; (2) the different outcomes neither need to be measured at the same time-points (in a longitudinal study), nor does one have to assume that the same number of repeated measurements is available for all outcomes; (3) a simple indication of the association between different outcomes is observed; (4) implementation can easily be done in standard statistical software packages (e.g., the SAS procedure MIXED for analyzing only continuous outcomes and the SAS procedure NLMIXED for the evaluation of discrete outcomes or mixed type outcomes), etc.

## 6.2 Basic Building Blocks for a Subject-Specific Interpretation

Consider two longitudinal outcomes $Y_{1ij}$ and $Y_{2ik}$, denoting the $j$th and $k$th measurement on the $i$th subject for outcome 1 and 2, respectively, ($i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_{1i}$, and $k = 1, 2, \ldots, n_{2i}$). With similar notation as before (Section 3.4.1), i.e., where both outcomes are modeled with the flexible modeling framework of Section 3.4, the ESPM for subject-specific interpretations can generally be expressed by the following three parts:

$$\eta_{1ij} = \mathbf{x}_{1ij}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{z}_{1ij}^{'} \cdot \mathbf{b}_{1i}, \tag{6.1}$$

$$\eta_{2ij} = \mathbf{x}_{2ij}^{'} \cdot \boldsymbol{\xi}_2 + \Lambda \cdot \mathbf{z}_{2ij}^{'} \cdot \mathbf{b}_{2i}, \tag{6.2}$$

$$\begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \end{bmatrix} \sim N\left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix} \right). \tag{6.3}$$

Here, a few comments are in place. First, the fixed-effects parameters $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ for the first outcome and second outcome, respectively, are kept separate because the type of outcome can be different, e.g., a time-to-event outcome for the first one and a continuous outcome for the second one. Second, the covariates in (6.1) and (6.2) need not be the same. Third, a (usually diagonal) re-scaling matrix $\Lambda$ is included in the GLMM part because of the difference of scale. If no scale difference is present between both outcomes (e.g., when two outcomes of the same type are modeled), the contribution of $\Lambda$ in (6.2) will vanish. Fourth, different random effect vectors are used for the different longitudinal profiles, i.e., $\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$. Both vectors are related by assuming that these random effects are bivariate normally distributed. Fifth, conditionally on the normally distributed random effects vector ($\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$), it is assumed that the two outcomes are independent. The common dependency between both outcomes is therefore specified by the correlated matrix $D_{12}$. If uncorrelated random effects vectors are assumed, the resulting model is equivalent to modeling the two outcomes separately by the flexible modeling framework of Section 3.4. Sixth, a different random-effects structure can be applied for both outcomes. For example, if a random-slope model is chosen for outcome 1 and a random-intercept model is used for outcome 2, the bivariate normal distribution (6.3) is expressed by

$$\begin{bmatrix} \begin{pmatrix} b_{1(0)i} \\ b_{1(1)i} \end{pmatrix} \\ b_{2(0)i} \end{bmatrix} \sim N\left( \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 0 \end{bmatrix}, \begin{bmatrix} \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} & \begin{pmatrix} d_{13} \\ d_{23} \end{pmatrix} \\ \begin{pmatrix} d_{13} & d_{23} \end{pmatrix} & d_{33} \end{bmatrix} \right). \tag{6.4}$$

Often, a general unstructured variance-covariance matrix will be assumed, but specific restrictions can be imposed as well. For example, assuming perfect correlations between elements in $\mathbf{b}_{1i}$ and elements in $\mathbf{b}_{2i}$, i.e., $d_{13} = d_{23} = 1$ in (6.3), would lead to a joint model in which some random effects are shared between the two outcomes. Evidently, this is equivalent to sharing components between $\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$.

Due to the independency between both outcomes, conditionally on the random effects, the likelihood function corresponding to the joint model is given by

$$L(\boldsymbol{\Xi}) = \prod_{i=1}^{N} \left[ \int_b \left( \underbrace{f_{1i}(\mathbf{y}_{1i} \mid \mathbf{b}_{1i}) \cdot f_{2i}(\mathbf{y}_{2i} \mid \mathbf{b}_{2i})}_{=f_i(\mathbf{y}_{1i}, \mathbf{y}_{2i} \mid \mathbf{b}_i)} \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i \right) \right], \tag{6.5}$$

in which $\boldsymbol{\Xi}$ is the vector of all parameters in the conditional distributions and the multivariate normal distribution of $\mathbf{b}_i$, and $f_{pi}(\mathbf{y}_{pi} \mid \mathbf{b}_{pi}) = \int_\theta f_{pi}(\mathbf{y}_{pi} \mid \boldsymbol{\theta}_{pi}, \mathbf{b}_{pi}) \cdot f_{pi}(\boldsymbol{\theta}_{pi}) \cdot d\boldsymbol{\theta}_{pi}$ $(p = 1, 2)$. Except for special cases (e.g., with linear models), the integral of (6.5) cannot be calculated analytically and numerical approaches (e.g., adaptive Gaussian quadrature) are needed.

Ivanova, Molenberghs and Verbeke (2013) used this terminology to model two ordinal outcomes jointly by two proportional odds mixed model (POMM), while Ghebretinsae et al (2013) obtained this methodology to model the tail length and tail intensity in the comet assay (Molenberghs and Verbeke, 2005) in a joint way with two random-intercept WGN models.

## 6.3 Applications

### 6.3.1 Case 1: Joint Hierarchical Multilevel Model for Longitudinal Continous and Repeated Time-to-Event Data

One of the most common practical situations is the collection of longitudinal continuous outcomes and repeated, time-to-event outcomes simultaneously. For example, Dendale et al. (2011) and Njagi et al. (2013a) describe a study in cardiology, in which researchers, through telemonitoring (a process through which patients are remotely monitored), not only repeatedly measured daily blood pressure, heart rate and weight from initially discharged chronic heart failure patients, but also recorded the time-to-rehospitalization. Time-to-rehospitalization in this case was a recurrent survival outcome, since a discharged patient could be rehospitalized more than once over time, while the repeatedly measured daily blood pressure, heart rate and weight of a patient each correspond to a longitudinal, continuous outcome.

Let $T_{ik}$ be the time-to-event outcome for the $k$th measurement in cluster $i$ $(k = 1, \ldots, p_i)$ and $Y_{ij}$ represent the continuous outcome for the $j$th measurement in cluster $i$ $(j = 1, \ldots, n_i)$. The proposed WGN model (4.5)–(4.8) with fixed parameters $\boldsymbol{\xi}_1$ is used to model $T_{ik}$. To avoid over-parametrization, the parameters of the Gamma random effects are kept constant, i.e., $\theta_{ik} \sim \text{Gamma}(\alpha, \beta)$, and the scale and shape parameters are assumed fixed between members of a cluster, i.e., $\lambda$ and $\rho$. Other choices can be made as well. Additionally, the gamma random effects are assumed independent, but can be relaxed (not been done here). For the continuous outcome $Y_{ij}$, the LMM of Section 3.3.1 is used with fixed parameters $\boldsymbol{\xi}_2$. The same terminology is used for this approach. Furthermore, a vector of scale factors $\mathbf{w}_{ik}^{'}$ is used in the time-to-event setting to resolve the issue of difference scaling. At last, shared normal random effects are considered, in such a generic way that $\mathbf{z}_{ij}^{'}$ and $\mathbf{w}_{ik}^{'}$ can be chosen such that some random effects are present in the Weibull predictor, others are only obtained in the linear predictor of the

normal outcome and the rest influence both outcomes. Hereby, both shared and correlated random effects are encompassed in the joint model.

Conditionally on the shared normal random effects, the LMM (for modeling the continuous outcome) and WGN model (for modeling the time-to-event outcome) are assumed independent. Therefore, the conditional ESPM distribution (on both the normal and gamma random effects) can be written as

$$f(\mathbf{y}_i, \mathbf{t}_i \mid \mathbf{b}_i, \boldsymbol{\theta}_i) = \prod_{k=1}^{p_i} \lambda \cdot \rho \cdot \theta_{ik} \cdot t_{ik}^{\rho-1} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} \cdot e^{-\lambda \cdot t_{ik}^{\rho} \cdot \theta_{ik} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i}} \cdot \frac{1}{(2 \cdot \pi)^{n_i/2} \cdot \mid \Sigma_i^{-1} \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2 - Z_i \cdot \mathbf{b}_i)^{'} \cdot \Sigma_i^{-1} \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2 - Z_i \cdot \mathbf{b}_i)}. \quad (6.6)$$

After some derivations (Appendix D.1.1), it can be shown that the marginal ESPM distribution obtains a closed form, i.e., formula $(D.6)$. While the marginal probabilities could be directly specified for estimation through maximum likelihood, the existence of infinite series, as can be seen from the expressions in Appendix D.1.1, may make the approach intractable. Therefore, due to the ease of analytical integration over the conjugate random effects, and the availability of software that can numerically integrate over normal random effects (e.g., adaptive Gaussian quadrature), partial marginalization (Molenberghs et al, 2007 and 2010) can be used. In this approach, only the expression for the joint distribution marginal over the conjugate but conditional on the normal random effects is needed, which in this case is given by

$$f(\mathbf{y}_i, \mathbf{t}_i \mid \mathbf{b}_i) = \frac{1}{(2 \cdot \pi)^{n_i/2} \cdot \mid \Sigma_i^{-1} \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2 - Z_i \cdot \mathbf{b}_i)^{'} \cdot \Sigma_i^{-1} \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2 - Z_i \cdot \mathbf{b}_i)} \cdot \prod_{k=1}^{p_i} \frac{\lambda \cdot \rho \cdot t_{ik}^{\rho-1} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} \cdot \alpha}{\left(\lambda \cdot t_{ik}^{\rho} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} + \frac{1}{\beta}\right)^{\alpha+1} \cdot \beta^{\alpha}}. \quad (6.7)$$

### 6.3.2 Case 2: Joint Hierarchical Multilevel Model for Repeated Binary Outcomes and Repeated Time-to-Event Data

A similar approach can be used to obtain an ESPM for repeated binary and repeated time-to-event outcomes. For similarity, $T_{ik}$ is defined as the repeated time-to-event outcome for the $k$th survival time in cluster $i$ ($k = 1, \ldots, p_i$). The repeated binary outcome is here denoted by $Y_{ij}$ (either 0 or 1), for the $j$th measurement in cluster $i$ ($j = 1, \ldots, n_i$). The WGN model from Section 6.3.1 is adopted to model the repeated time-to-event outcome $T_{ik}$. For the repeated binary outcome $Y_{ij}$, the probit-beta-normal (PBN) model will be used. A general formulation of the PBN is given by

$$Y_{ij} \mid \mathbf{b}_i, \theta_{ij} \sim \text{Bernoulli}(\pi_{ij}), \quad (6.8)$$

$$\pi_{ij} = \theta_{ij} \cdot \Phi(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}_2 + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i), \quad (6.9)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \quad (6.10)$$

$$\theta_{ij} \sim \text{Gamma}(\alpha_1, \beta_1). \quad (6.11)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution and the parameters of the Gamma distribution, i.e., $\alpha_1$ and $\beta_1$, are held fixed to avoid over-parametrization. Molenberghs et al (2010) pointed out that closed-form expressions exist for the mean, variance and marginal distribution of the PBN model, making this methodology mathematically convenient to use. Even though the logit-beta-normal (LBN) model does not possess this property, Johnson and Kotz (1970) showed that

$$\frac{e^y}{1+e^y} \approx \Phi_1(c \cdot y), \tag{6.12}$$

with $c = (16 \cdot \sqrt{3})/(15 \cdot \phi)$, indicating that the LBN model can be used as mathematically convenient alternative for modeling the repeated binary outcome $Y_{ij}$. Additionally, shared normal random effects are considered in the same generic way as Section 6.3.1, while, conditionally on these shared normal random effects, both models are assumed independently of each other.

By integrating out the beta random effects from the PBN model (6.8)–(6.11), it can easily be shown that

$$f(y_{ij} \mid \mathbf{b}_i) = \frac{1}{\alpha_1+\beta_1} \cdot (K_{ij} \cdot \alpha_1)^{y_{ij}} \cdot [(1-K_{ij}) \cdot \alpha_1 + \beta_1]^{1-y_{ij}}, \tag{6.13}$$

with

$$K_{ij} = \Phi(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}_2 + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i). \tag{6.14}$$

The joint conditional ESPM can then be formulated by

$$\begin{aligned} f(\mathbf{y}_i, \mathbf{t}_i \mid \mathbf{b}_i, \boldsymbol{\theta}_i) = &\prod_{k=1}^{p_i} \lambda \cdot \rho \cdot \theta_{ik} \cdot t_{ik}^{\rho-1} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} \cdot e^{-\lambda \cdot t_{ik}^{\rho} \cdot \theta_{ik} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i}} \\ &\cdot \prod_{j=1}^{n_i} \frac{1}{\alpha_1+\beta_1} \cdot (K_{ij} \cdot \alpha_1)^{y_{ij}} \cdot [(1-K_{ij}) \cdot \alpha_1 + \beta_1]^{1-y_{ij}}. \end{aligned} \tag{6.15}$$

The conditioning here is only on the gamma and normal random effects, given that the beta random effects in the PBN model (6.8)–(6.11) has been integrated out.

Similar to the first case (Section 6.3.1), closed-form expressions can be derived for the marginal ESPM distribution (depending on the situation). For example, when the binary outcome repeatedly measures 1, i.e., $\mathbf{y}_i = (1, 1, \ldots, 1) = \mathbf{1}$, a closed-expression is derived for the marginal ESPM distribution (Appendix D.1.2). Due to the existence of infinity series in this marginal ESPM distribution, i.e., formula $(D.11)$, partial marginalization can be chosen to conduct inferences (similar to Section 6.3.1). Integrating out the Gamma random effects in (6.15), lead to the joint distribution conditional on the normal random effects:

$$f(\mathbf{y}_i, \mathbf{t}_i \mid \mathbf{b}_i) = \prod_{j=1}^{n_i} \frac{1}{\alpha_1 + \beta_1} \cdot (K_{ij} \cdot \alpha_1)^{y_{ij}} \cdot [(1 - K_{ij}) \cdot \alpha_1 + \beta_1]^{1-y_{ij}}$$
$$\cdot \prod_{k=1}^{p_i} \frac{\lambda \cdot \rho \cdot t_{ik}^{\rho-1} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} \cdot \alpha_1}{\left(\lambda \cdot t_{ik}^{\rho} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i} + \frac{1}{\beta_1}\right)^{\alpha_1+1} \cdot \beta_1^{\alpha_1}}. \tag{6.16}$$

## 6.4 Basic Building Blocks for Population-Averaged Interpretations

Analogous to Section 6.2, a similar approach can be followed for population-averaged interpretations. Unlike the subject-specific methodology, a few changes need to be made for (6.1) and (6.3) (according to Section 5.1).

Assuming the same terminology as before (Section 6.2), i.e., two longitudinal outcomes $Y_{1ij}$ and $Y_{2ik}$, and by modeling these with the COMMM of Section 5.1, the ESPM for population-averaged interpretations can generally be expressed by the following three parts:

$$\eta_{1ij} = \Delta_{1ij} + \mathbf{z}_{1ij}^{'} \cdot \mathbf{b}_{1i}, \tag{6.17}$$
$$\eta_{2ij} = \Delta_{2ij} + \Lambda \cdot \mathbf{z}_{2ij}^{'} \cdot \mathbf{b}_{2i}, \tag{6.18}$$
$$\begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \end{bmatrix} \sim N\left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix} \right). \tag{6.19}$$

Here, $\Delta_{1ij}$ and $\Delta_{2ij}$ represent the connector function between (5.2) and (5.3) for outcome $Y_{1ij}$ and $Y_{2ik}$, respectively. While the same comments hold as for the subject-specific approach (Section 6.2), a population-averaged interpretation is now present for the fixed-effects parameters $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ of the first outcome and second outcome, respectively.

Efendi et al (2012) used this approach to supplement the work of Njagi et al. (2012) by proposing a marginalized joint model for longitudinal continuous and repeated time-to-event outcomes, as well as a marginalized joint model for bivariate repeated time-to-event outcomes. While the former one was used for testing a joint effect of heart rhythm on repeated time-to-hospitalization as well as on the longitudinal heart rate of chronic heart failure (CHF) patients, the latter one was conducted on the comet assay for modeling the percentage of tail intensity and tail moment together.

## 6.5 Applications

### 6.5.1 Case 3: Joint Marginalized Multilevel Model for Bivariate, Repeated Time-to-Event Data

Nowadays, extensive research has been done on the ESPM for bivariate, repeated time-to-event data. For example, Ghebretinsae et al (2012) used the ESPM framework to

model the tail length and tail intensity in the comet assay data. In this setting, the so-called Weibull-Gamma-Multivariate-Normal model (WGMNM) for subject-specific interpretations was proposed, with the use of random-intercepts in the linear parts. Closed form expressions and derivations have been given for the correlation between both endpoints and intraclass correlation (ICC). Efendi et al (2012) redefined this methodology for population-averaged interpretations (Section 6.4).

Let $Y_{1ij}$ be the $j$th survival time in cluster $i$ ($j = 1, ..., p_i$) of outcome 1 and $Y_{2ik}$ represent the $k$th survival time in cluster $i$ ($k = 1, ..., n_i$) of outcome 2. The marginalized WGN model (Section 5.1) is used to model both outcomes $Y_{1ij}$ and $Y_{2ij}$ separately, conditional upon the normal random effects. Equivalent as before (Section 6.3.1), the parameters of the Gamma random effects are kept constant, i.e., $\theta_{1ij} \sim \text{Gamma}(\alpha_1, \beta_1)$ and $\theta_{2ik} \sim \text{Gamma}(\alpha_2, \beta_2)$, and the scale and shape parameters are assumed fixed between members of a cluster, i.e., $\lambda_1$, $\rho_1$, $\lambda_2$ and $\rho_2$, in order to avoid over-parametrization. Furthermore, the gamma random effects are assumed to be independent of each other, and also independent from the normal random effects. Other choices can be made as well. At last, both shared and correlated random effects are encompassed in the joint model.

The conditional ESPM density can be expressed as:

$$f(\mathbf{y}_{1i}, \mathbf{y}_{2i} \mid \mathbf{b}_{1i}, \mathbf{b}_{2i}, \boldsymbol{\theta}_{1i}, \boldsymbol{\theta}_{2i}) = \prod_{j=1}^{p_i} \lambda_1 \cdot \rho_1 \cdot \theta_{1ij} \cdot y_{1ij}^{\rho_1 - 1} \cdot e^{\Delta_{1ij} + \mathbf{z}_{1ij}' \cdot \mathbf{b}_{1i}} \cdot e^{-\lambda_1 \cdot y_{1ij}^{\rho_1} \cdot \theta_{1ij} \cdot e^{\Delta_{1ij} + \mathbf{z}_{1ij}' \cdot \mathbf{b}_{1i}}}$$
$$\cdot \prod_{k=1}^{n_i} \lambda_2 \cdot \rho_2 \cdot \theta_{2ik} \cdot y_{2ik}^{\rho_2 - 1} \cdot e^{\Delta_{2ij} + \mathbf{z}_{2ik}' \cdot \mathbf{b}_{2i}} \cdot e^{-\lambda_2 \cdot y_{2ik}^{\rho_2} \cdot \theta_{2ik} \cdot e^{\Delta_{2ik} + \mathbf{z}_{2ik}' \cdot \mathbf{b}_{2i}}}. \tag{6.20}$$

Due to the mutually independency of the gamma distributed components, the derivation of the marginal ESPM distribution simplifies to:

$$\begin{aligned} f(\mathbf{y}_{1i}, \mathbf{y}_{2i}) = & \sum_{(m_{11}, ..., m_{1n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_{1j}}}{m_{1j}!} \cdot \frac{\Gamma(\alpha_1 + m_{1j} + 1) \cdot \beta_1^{m_{1j}+1}}{\Gamma(\alpha_1)} \cdot (\lambda_1^{m_{1j}+1} \cdot \rho_1 \cdot y_{1ij}^{(m_{1j}+1)\cdot\rho_1 - 1})^{\delta_{1i}} \\ & \cdot (\lambda_1^{m_{1j}} \cdot y_{1ij}^{\rho_1 \cdot m_{1j}})^{1-\delta_{1i}} \cdot e^{(m_{1j}+\delta_{1i})\cdot\Delta_{1ij} + \frac{1}{2}\cdot\mathbf{z}_{1ij}'\cdot D_{11}\cdot\mathbf{z}_{1ij}\cdot(m_{1j}+\delta_{1i})^2} \cdot \sum_{(m_{21}, ..., m_{2n_i})} \prod_{k=1}^{p_i} \frac{(-1)^{m_{2j}}}{m_{2j}!} \\ & \cdot \frac{\Gamma(\alpha_2 + m_{2j} + 1) \cdot \beta_2^{m_{2j}+1}}{\Gamma(\alpha_2)} \cdot (\lambda_2^{m_{2j}+1} \cdot \rho_2 \cdot y_{2ij}^{(m_{2j}+1)\cdot\rho_2 - 1})^{\delta_{2i}} \cdot (\lambda_2^{m_{2j}} \cdot y_{2ij}^{\rho_2 \cdot m_{2j}})^{(1-\delta_{2i})} \\ & \cdot e^{(m_{2j}+\delta_{2i})\cdot\Delta_{2ij} + \frac{1}{2}\cdot\mathbf{z}_{2ij}'\cdot D_{22}\cdot\mathbf{z}_{2ij}\cdot\{(m_{2j}+1)^{2\cdot\delta_{2i}} + [m_{2j}^2\cdot(1-D_{12}^2) + D_{12}^2\cdot m_{2j}]^{(1-\delta_{2i})}\}} \\ & \cdot e^{(\mathbf{z}_{1ij}'\cdot|D_{11}|^{1/2}\cdot\mathbf{z}_{1ij})\cdot(\mathbf{z}_{2ij}'\cdot|D_{22}|^{1/2}\cdot\mathbf{z}_{2ij})\cdot D_{12}\cdot(m_{1j}+\delta_{1i})\cdot(m_{2j}+\delta_{2i})}. \end{aligned} \tag{6.21}$$

Due to the existence of infinite series in this marginal ESPM distribution, i.e., formula (6.21), partial marginalization can be chosen to conduct inferences (similar to Section 6.3.1 and Section 6.3.2).

Even though the association between both outcomes is captured by the bivariate normal random effects, i.e., formula (6.19), the correlation between the two random effects $\mathbf{b}_{1i}$

and $\mathbf{b}_{2i}$ is not necessarily equal to the correlation between the two responses $\mathbf{y}_{1i}$ and $\mathbf{y}_{2i}$. Furthermore, a significant correlation at the cluster level does not necessarily imply a significant correlation between the two responses taken from the same cell. Therefore, special attention is given to the expressions and derivations of (1) the correlation between two measurements from the same outcome for a single cluster, also called the intraclass correlation (ICC), and (2) the correlation between the two outcomes of the same cluster measurement.

In this section, explicit expressions and derivations are presented for the random-intercept approach, where censoring is not present in both outcomes (Appendix D.2). Mathematically, the linear part for the two outcomes $Y_{1ij}$ and $Y_{2ik}$ and bivariate normal distribution for the two random effects of this approach are expressed by

$$\eta_{1ij} = \Delta_{1ij} + b_{1i}, \tag{6.22}$$

$$\eta_{2ij} = \Delta_{2ij} + b_{2i}, \tag{6.23}$$

$$\begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} d_1^2 & r \cdot d_1 \cdot d_2 \\ r \cdot d_1 \cdot d_2 & d_2^2 \end{bmatrix} \right), \tag{6.24}$$

with

$$\Delta_{1ij} = -\log(\alpha_1 \cdot \beta_1) + \mathbf{x}_{1ij}^{'} \boldsymbol{\xi}_1^m - \frac{d_1^2}{2},$$

$$\Delta_{2ij} = -\log(\alpha_2 \cdot \beta_2) + \mathbf{x}_{2ij}^{'} \boldsymbol{\xi}_2^m - \frac{d_2^2}{2}.$$

Here, $d_1^2$ and $d_2^2$ represent the variances of the cluster-specific random intercepts $b_{1i}$ and $b_{2i}$, respectively, while $r$ expresses the correlation between them. For reasons of identifiability (Duchateau and Janssen, 2007), $\alpha_v \cdot \beta_v = 1$ is chosen ($v = 1, 2$). Focus is given on the relationship between the correlation among the outcomes and the correlation among the random effects.

While Fitzmaurice et al (2009) showed that, for a joint model based on two linear mixed model, the bivariate correlation between the two outcomes can be expressed by $r \cdot \sqrt{\text{ICC}_1} \cdot \sqrt{\text{ICC}_2}$, where $\text{ICC}_1$ and $\text{ICC}_2$ denotes the intraclass correlation for outcome 1 and 2, respectively, Ghebretinsae et al (2012) extended this approach for (6.22) – (6.24). The remainder of this section is devoted on this approach.

1. *The correlation between two measurements from the same outcome for a single cluster*

   The intraclass correlation (ICC) equals to

$$\text{Corr}(Y_{lij}, Y_{lik}) = \frac{e^{d_l^2/\rho_l^2} - 1}{\frac{2 \cdot \rho_l \cdot B(\alpha_l - \frac{2}{\rho_l}, \frac{2}{\rho_l}) \cdot e^{d_l^2/\rho_l^2}}{B(\alpha_l - \frac{1}{\rho_l}, \frac{1}{\rho_l})^2} - 1}, \tag{6.25}$$

with $l = 1, 2$ and $B(\cdot, \cdot)$ the beta function. Moreover, $\alpha_l$, $\rho_l$ and $d_l^2$ represent the conventional shape parameter of the Gamma random effects distribution, shape parameter of the Weibull distribution, and random-effects variance of outcome $l$, respectively.

2. *The correlation between the two outcomes of the same cluster measurement*

The correlation between the two outcomes of the same cluster measurement is expressed by

$$\begin{aligned}
\text{Corr}(Y_{1ij}, Y_{2ij}) &= \left(e^{\frac{r \cdot d_1 \cdot d_2}{\rho_1 \cdot \rho_2}} - 1\right) \\
&\quad \times \frac{B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1}\right)}{\left[2 \cdot \rho_1 \cdot B\left(\alpha_1 - \frac{2}{\rho_l}, \frac{2}{\rho_l}\right) \cdot e^{d_1^2/\rho_1^2} - B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_l}\right)^2\right]^{1/2}} \\
&\quad \times \frac{B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)}{\left[2 \cdot \rho_2 \cdot B\left(\alpha_2 - \frac{2}{\rho_2}, \frac{2}{\rho_2}\right) \cdot e^{d_2^2/\rho_2^2} - B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)^2\right]^{1/2}} \\
&= \frac{\left(e^{\frac{r \cdot d_1 \cdot d_2}{\rho_1 \cdot \rho_2}} - 1\right)}{\left(e^{d_1^2/\rho_1^2} - 1\right)^{1/2} \cdot \left(e^{d_2^2/\rho_2^2} - 1\right)^{1/2}} \cdot \sqrt{\text{ICC}_1} \cdot \sqrt{\text{ICC}_2}. \qquad (6.26)
\end{aligned}$$

Thus, the correlation between the two endpoints is proportional to the correlation between the two random effects, with the same sign. In other words, the correlation between the two endpoints is entirely induced by the correlation between the two random effects. Additionally, this correlation also depends on the Weibull shape parameters $\rho_1$ and $\rho_2$, and variance and covariance elements of the bivariate normal distribution for the random effects, i.e., $d_1^2$, $d_2^2$ and $r \cdot d_1 \cdot d_2$.

### 6.5.2 Case 4: Joint Marginalized Multilevel Model for capturing Informative Censoring Mechanisms in Overdispersed, Repeated Time-to-Event Data

Most statistical methods for censored time-to-event data almost invariably assume that there is no dependency present between the lifetime and censoring mechanisms, an assumption that is often doubtful in practice. Standard modeling techniques like the Cox proportional hazard model (Cox, 1972) rely on this assumption and needed to be handled with careful attention. In clinical trials, for example, it often occurs that patients withdraw from the study. A possible reason is the better conditional state of the patient at that moment, such that no further medical attention is needed. In this case, the event that was proceeded before the censoring may have a significant effect, and may increase the expected remaining lifetime. Lagakos (1979) formulate an amount of examples where the assumption of non-informative censoring is doubtful.

Until now, non-informative censoring was assumed for the hierarchical and marginalized WGN model of Section 4.1 and Section 5.1, respectively. Dealing with informative censoring substantially increases the complexity of modeling. In the past, several methods have been described to deal with the problem of informative censoring. These include imputation techniques for missing data, sensitivity analyses to mimic best and worst-case scenarios and use of the drop-out event as a study end-point (Shih, 2002). In this section, a new approach is developed, by defining an ESPM that captures the occurrence of informative censoring in an overdispersed, repeated time-to-event data structure. Similar as before, focus will be given on right censoring.

Let $T_{ij}$ be the $j$th (known) lifetime in cluster $i$, $C_{ij}$ represents the $j$th censored time for cluster $i$ and $I_{ij}$ denotes the $j$th censored indication in cluster $i$ ($i = 1, \ldots, N$; $j = 1, \ldots, n_i$), which equals to 0 if $T_{ij} \leq C_{ij}$ and 1 if $T_{ij} > C_{ij}$. The $j$th time-to-event outcome $Y_{ij}$ for cluster $i$ can then simply be formulated by

$$Y_{ij} = \min(T_{ij}, C_{ij}) = \begin{cases} T_{ij} & \text{, if } T_{ij} \leq C_{ij} \\ C_{ij} & \text{, if } T_{ij} > C_{ij} \end{cases}. \tag{6.27}$$

The marginalized WGN model from Section 5.1 is used to model outcome $Y_{ij}$. For the repeated binary outcome $I_{ij}$, an extended version of a marginalized specified logistic-normal model (Heagerty, 1999; Heagerty and Zeger, 2000) will be used, that allows different link functions for the marginal and conditional specification (Griswold and Zeger, 2004). Here, a logit link function is chosen for the marginal model and a probit link is used for the conditional model. In this case, the odds ratio interpretation of the marginal parameters retains while taking advantage of the computational ease emanating from the probit-normal relationship. A closed form expression is derived for the connector function $\Delta_{2ij}$ (of $I_{ij}$), formulated by

$$\Delta_{2ij} = \sqrt{1 + \mathbf{z}_{2ij}^{'} \cdot D_{22} \cdot \mathbf{z}_{2ij}} \cdot \Phi^{-1}[\text{expit}(\mathbf{x}_{2ij}^{'} \cdot \boldsymbol{\xi}_2^m)], \tag{6.28}$$

with $\mathbf{x}_{2ij}$ and $\mathbf{z}_{2ij}$ the $p$- and $q$-dimensional vectors of known covariate values, $\boldsymbol{\xi}_2^m$ the marginal fixed-effects parameters and $D_{22}$ the variance-covariance matrix of the random effect $\mathbf{b}_{2i}$ for outcome $I_{ij}$. Additionally, $\Phi^{-1}(\cdot)$ denotes the conventional probit link function. A general formulation of this model is expressed by

$$\text{logit}(\mu_{2ij}^m) = \mathbf{x}_{2ij}^{'} \cdot \boldsymbol{\xi}_2^m, \tag{6.29}$$

$$\Phi^{-1}(\mu_{2ij}^c) = \sqrt{1 + \mathbf{z}_{2ij}^{'} \cdot D_{22} \cdot \mathbf{z}_{2ij}} \cdot \Phi^{-1}[\text{expit}(\mathbf{x}_{2ij}^{'} \cdot \boldsymbol{\xi}_2^m)] + \mathbf{z}_{2ij}^{'} \cdot \mathbf{b}_{2i}, \tag{6.30}$$

with $\mu_{2ij}^m = E(I_{ij})$, i.e., the marginal mean of $I_{ij}$, and $\mu_{2ij}^c = E(I_{ij} \mid \mathbf{b}_i)$, i.e., the conditional mean of $I_{ij}$. This model, i.e., (6.29)–(6.30), enjoys a likelihood basis and allows derivation of the full probability distribution for the response (Fitzmaurice and Laird, 1993; Molenberghs and Lesaffre, 1994). Furthermore, Iddi and Molenberghs (2012) pointed out that the model produces valid inferences when data are missing at random (MAR). A comprehensive discussion of this missing value taxonomy will be given in Chapter 7.

Conditionally on the shared normal random effects, the marginalized WGN model (for modeling time-to-event outcome $Y_{ij}$) and model (6.29)–(6.30) (for modeling the binary

outcome $I_{ij}$) are assumed independent. Similar as before (Section 6.3.1), both shared and correlated random effects are encompassed in the model. A general formulation of this framework is given by

$$\mu^c_{1ij} = \theta_{1ij} \cdot exp\left[\mathbf{x}'_{1ij} \cdot \boldsymbol{\xi}^m_1 - (\mathbf{z}'_{1ij} \cdot D_{11} \cdot \mathbf{z}_{1ij})/2 - log(\alpha_1 \cdot \beta_1) + \mathbf{z}'_{1ij} \cdot \mathbf{b}_{1i}\right], \tag{6.31}$$

$$\mu^c_{2ij} = \Phi\left\{\sqrt{1 + \mathbf{z}'_{2ij} \cdot D_{22} \cdot \mathbf{z}_{2ij}} \cdot \Phi^{-1}[expit(\mathbf{x}'_{2ij} \cdot \boldsymbol{\xi}^m_2)] + \Lambda \cdot \mathbf{z}'_{2ij} \cdot \mathbf{b}_{2i}\right\}, \tag{6.32}$$

$$\begin{bmatrix}\mathbf{b}_{1i}\\ \mathbf{b}_{2i}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}D_{11} & D_{12}\\ D_{12} & D_{22}\end{bmatrix}\right), \tag{6.33}$$

with $\mu^c_{1ij} = E(Y_{ij} \mid \mathbf{b}_{1i})$, i.e., the conditional mean of time-to-event outcome $Y_{ij}$, and $\mu^c_{2ij} = E(I_{ij} \mid \mathbf{b}_i)$, i.e., the conditional mean of binary outcome $I_{ij}$. Similar as before (Section 6.5.1), the same assumptions are made. Other choices can be made as well.

This technique enables us (1) to capture the informative censoring by the correlated normal random effects and (2) obtain parameter estimates that have a population-averaged interpretation for both outcomes. Estimation strategies such as maximum likelihood estimation with partial marginalization (Section 4.2.1), pairwise likelihood (Section 4.2.2) and Bayesian estimation (Section 4.2.3) can be performed.

# Chapter 7

# A Characterization of Missingness at Random in a Generalized Shared-Parameter Joint Modeling Framework for Longitudinal and Time-to-Event Data

Statistical modeling is often connected with the complexity of missing data. In this chapter, a conceptual correspondence is described between the missing data setting, and joint modeling of longitudinal and time-to-event outcomes, according to Njagi et al (2013c). Hereby, a formulation is presented of an extended shared random effects joint model, similar to that of Creemers et al. (2011) in the context of longitudinal data subject to missing observations. A characterizations of missing at random (MAR) is provided within the missing data setting. While an additional complexity arises in the joint longitudinal and time-to-event setting, i.e., coarsening (see Chapter 10 for a more profound discussion), an avenue for sensitivity analyses is considered.

The chapter is organized as follows. In Section 7.2, a brief review is given on missing data. In particular, attention is provided on the different modeling frameworks and the characterization of MAR in each of these frameworks. The generalized shared-parameter modeling (GSPM) framework of Creemers et al. (2011), and its MAR characterization, are discussed as well. In Section 7.3, three main scenario's are examined to illustrate the correspondence between joint modeling of longitudinal and time-to-event data, and missing data. First, the extended framework of Njagi et al (2013c) is defined, alongside with the corresponding MAR characterization. At last, in Section 7.4, an avenue for sensitivity analyes is provided for this setting.

## 7.1 Introduction

Missing data often occur in practical situations, and need to be taken into account when proceeding with statistical inference. In the past, three main models have been developed in the missing data setting, and will be considered as starting point in this chapter.

When missingness is present in the data, models are often catalogued in one of the following three modeling frameworks: (1) selection models (SEM), (2) pattern-mixture models (PMM) and (3) shared-parameter models (SPM). While the former two approaches are defined by different factorizations of the joint distribution of the data and the missing value processes, the latter one is defined by assuming that both data and the missing value process depend on latent variables, conditional upon which independence is assumed.

Within the SEM, Rubin (1976) classified the missing value processes into several parts. In a frequentist approach (Section 3.5.1), with outcomes only missing, the classification can be expressed in three parts: (1) Missing completely at random (MCAR), meaning that, conditional upon covariates, the missing value mechanism does not depend on outcomes, (2) missing at random (MAR), meaning that, conditional on covariates and observed outcomes, the missing value mechanism does not further depend on missing outcomes, and (3) missing not at random (MNAR), implying that, conditional on covariates and observed outcomes, the missing value mechanism does depend on unobserved outcomes. Since Rubin (1976) developed these processes within the SEM, Molenberghs et al (1998) and Creemers et al (2011) transposed this taxonomy to the PMM and SPM framework, respectively.

For the particular case of longitudinal data with dropout, Molenberghs et al (1998) derived a set of so-called identifying restrictions, to identify the model for the missing measurements given the observed ones within a missing-data pattern, consistent with MAR. Furthermore, Molenberghs et al. (2007) showed that for every MNAR model, there is an MAR counterpart that produces exactly the same fit to the observed data in the PMM framework. Hence, the original model and its MAR counterpart cannot be distinguished from one another. Creemers et al (2011), on the other hand, showed that a similar connection can be made with the MAR counterpart in the sense of Molenberghs et al (2007), but now in the SPM framework.

Since this thesis partly focuses on the joint modeling of longitudinal and time-to-event data, questions arises whether a characterization of MAR is present in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data. Differentiating from the other provided research, the joint longitudinal and time-to-event setting is slightly different, given that a time-to-event is also collected. Therefore, the objective may be three-fold, i.e., (1) to study the time-to-event outcome, accounting for the longitudinal covariate, (2) to study the longitudinal outcome, accounting for possibly non-random drop-out caused by the occurrence of events and (3) to examine the association structure between both outcomes (Tsiatis and Davidian, 2004).

Focuses on the most common one of the three, i.e., the first one, Njagi et al (2013c) noted that the objective is usually achieved within the SPM framework, where, conditional on a shared latent structure, e.g., a normal random effect, a sub-model for the time-to-event outcome is fitted independently upon the longitudinal process. Both outcomes are linked by the shared latent structure (Tsiatis and Davidian, 2004). However, attention needs to be given here. For example, Njagi et al (2013c) mentioned that the longitudinal covariate may be measured with error, its values are likewise only available at the specific time points at that the patient appears at the clinic for longitudinal measurements, and the time-to-event may also be censored.

Although the occurrence of pre-defined error determination in the longitudinal process is often problematic in making correct inferences about the time-to-event outcome, the joint density incorporates both censoring and the visiting and measurement probabilities (Tsiatis and Davidian, 2004), where the visiting probabilities represent the process which generate the time points at which measurements are available (Rizopoulos, 2012a). In order to identify the relationship of interest, it is assumed that, under likelihood inference, the probabilities of censoring and visiting can depend on past visit times and longitudinal measurements, but not further on the future longitudinal measurements and event time (Tsiatis and Davidian, 2004). From these statements, Njagi et al (2013c) pointed out that the above mentioned conditions mirror the MAR assumption mentioned earlier. Since the assumptions are unverifiable based on the data, equivalent to the unverifiable assumptions about the missing value mechanism in the missing data setting, sensitivity analysis is a recurring theme here (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005).

From the discussion above, it can be said that a strong connection between the joint longitudinal and time-to-event and missing data setting is present. To stay in line with the followed approach of Njagi et al (2013c), a slightly different perspective on joint models is taken than is prevalent in the literature. However, conceptually, the two settings actually correspond. By using the shared random effects joint model of Creemers et al. (2011) as starting point in the context of longitudinal data subject to missing data, Njagi et al (2013c) formulated an extended shared random effects joint model for longitudinal and time-to-event setting, where an added layer of complexity, i.e., coarsening, meaning that the actually observed data is less detailed than what is planned, can be present in the data. For example, the time-to-event outcome can be censored, and/or the longitudinal profiles may be incomplete. In statistical terminology, coarsening can be seen as one of the two basic aspects of enriched data. A full discussion of this terminology can be found in Chapter 10.

Within the extended framework of Njagi et al (2013c), a characterization of MAR is provided, consistent to the one in the missing data setting. Since the missing data setting empowers a lot of terminology, a brief discussion is first provided!

## 7.2 General discussion on missing data terminology

### 7.2.1 Notation and Modeling Frameworks

Similar as before (Section 3.2), let $Y_{ij}$ represent the $j$th outcome measured for cluster (subject) $i$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$. To account for the missing data process in the data, assume $R_{ij}$ to be a missingness indicator, which indicates 1 if $Y_{ij}$ is observed, and 0 otherwise. Thus, vectors $\mathbf{Y}_i$ and $\mathbf{R}_i$ represent the measurement and missingness process for subject $i$, respectively. Moreover, vector $\mathbf{Y}_i$ is partitioned into two vectors $\mathbf{Y}_i^0$ and $\mathbf{Y}_i^m$, containing the observed and unobserved components, respectively, and $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ be the parameter vectors for the measurement and missingness processes, respectively. Suppressing the covariate $\mathbf{x}_i$ from notation and expressing $f(\mathbf{y}_i, \mathbf{r}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi})$ as the full data density, a general discussion can be made for the missing data modeling frameworks.

In the missing data setting, models are often catalogued in three modeling frameworks. The first one, known as the selection model (SEM), starts from the factorization $f(\mathbf{y}_i, \mathbf{r}_i \mid \boldsymbol{\theta}, \psi) = f(\mathbf{y}_i \mid \boldsymbol{\theta}) \cdot f(\mathbf{r}_i \mid \mathbf{y}_i, \psi)$. In the second one, also called the pattern-mixture model (PMM), a slightly different factorization scheme is used, i.e., $f(\mathbf{y}_i, \mathbf{r}_i \mid \boldsymbol{\theta}, \psi) = f(\mathbf{y}_i \mid \mathbf{r}_i, \boldsymbol{\theta}) \cdot f(\mathbf{r}_i \mid \psi)$. For the last one, called the shared-parameter model (SPM), a vector of shared latent variables $\mathbf{b}_i$ is present, conditional upon which independence of the measurement and missingness processes is assumed, i.e.,

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \boldsymbol{\theta}, \psi) = \int f(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\theta}) \cdot f(\mathbf{r}_i \mid \mathbf{b}_i, \psi) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i. \tag{7.1}$$

### 7.2.2 Characterization of Missing at Random

Within each discussed framework, MAR can be defined. Under the SEM, the missing data mechanism is defined as MAR if $f(\mathbf{r}_i \mid \mathbf{y}_i, \psi) = f(\mathbf{r}_i \mid \mathbf{y}_i^0, \psi)$. In the case of PMM (Molenberghs et al, 1998), missingness is defined MAR if

$$f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{r}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \boldsymbol{\theta}). \tag{7.2}$$

In other words, MAR in PMM can be seen in such a way that the unobserved outcomes can be predicted from the observed outcomes and covariates, without further reference to the missingness mechanism.

In the SPM framework, i.e., (7.1), Creemers et al (2011) and Njagi et al (2013c) pointed out that MAR cannot hold without reducing to MCAR, where $\mathbf{b}_i$ drops from at least one of the factors in the integrand of (7.1). Therefore, Creemers et al (2011) generalized the SPM by expanding the random-effects structure, and called it the generalized SPM (GSPM) framework:

$$\begin{aligned} f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{q}_i) = f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{l}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) \\ \cdot f(\mathbf{r}_i \mid \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i). \end{aligned} \tag{7.3}$$

A few comments are in place here. First, a more general random-effects structure is present, where the random effects $\mathbf{g}_i$ are shared among all processes, $\mathbf{h}_i$, $\mathbf{j}_i$ and $\mathbf{k}_i$ are shared between two processes only, and $\mathbf{l}_i$, $\mathbf{m}_i$ and $\mathbf{q}_i$ are specific to one process. Second, a general condition can be made for MAR. Creemers et al (2011) denoted that GSPM is MAR of and only if

$$\begin{aligned} \frac{\int f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i) \cdot f(\mathbf{r}_i \mid \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}{\int f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{j}_i) \cdot f(\mathbf{r}_i \mid \mathbf{g}_i, \mathbf{j}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i} = \\ \frac{\int f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{h}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{g}_i, \mathbf{h}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}{f(\mathbf{y}_i^0)}. \end{aligned} \tag{7.4}$$

Third, the following convenient proper sub-class of GSPM (7.3) exists that satisfies MAR:

$$\begin{aligned} f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{q}_i) &= f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{j}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{q}_i) \\ &= f(\mathbf{y}_i^0 \mid \mathbf{j}_i, \mathbf{l}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{m}_i) \cdot f(\mathbf{r}_i \mid \mathbf{j}_i, \mathbf{q}_i). \end{aligned} \tag{7.5}$$

## 7.3 Joint Modeling of longitudinal and time-to-event data

While Creemers et al (2011) limited their discussion of GSPM in the context of of longitudinal data subject to missing data, Njagi et al (2013c) extended this to the joint modeling setting of longitudinal and time-to-event data. Here, three scenarios have been used to illustrate and motivate the correspondence between the joint modeling of longitudinal and time-to-event data and that of missing data. Since other scenarios are possible as well, this thesis only considers to these three scenarios.

The following three scenarios are used:

1. For subjects who drop out before the planned end of the study, longitudinal information prior to drop-out is observed, as well as the censoring time. Neither the latter part of the longitudinal sequence nor the survival time is present for these subjects.

2. For subjects who experience the event within the study period, such that the event censors the longitudinal sequence, longitudinal information prior to the event is observed, as well as the survival time. Longitudinal data after the event, as well as the censoring time, are unobserved for these subjects.

3. For subjects who reach the end of the study without experiencing the event, full longitudinal information as well as the censoring time are observed. The survival time remains unobserved for these subjects.

These three possible outcomes imply that the joint modeling of longitudinal and time-to-event always entails two parts, i.e., (1) a part that is observed and (2) a part that is unobserved. Furthermore, the mechanism that causes the coarsening, consisting of the union of the missingness mechanism in the longitudinal outcome, and a certain choice mechanism, related to the time- to-event outcome, which determines whether either the event time or censoring time is observed, must also need to be taken into account. As can be examined, complexity exponentially grows when dealing with more advanced settings.

Since the notation of Section 7.2.1 does not fully describe the joint modeling setting of Njagi et al (2013c), additionally notations are added. First, let $T_i$ and $C_i$ represent the survival and censoring times, respectively. Second, let $D_i^0 = \min(T_i, C_i)$ and $D_i^m = \max(T_i, C_i)$. Third, a vector of missingness indicators is introduced, i.e., $\mathbf{R}_i^* = (\mathbf{R}_i', W_i)'$, where $W_i = 1$ if the survival time is observed and 0 otherwise. With these additionally variables, the full set of (stochastic) defined components equals

$$\mathbf{Q}_i = \left(\mathbf{Y}_i^{0\prime}, \mathbf{Y}_i^{m\prime}, D_i^0, D_i^m, \mathbf{R}_i^{*\prime}\right)' = \left(\mathbf{Z}_i^{0\prime}, \mathbf{Z}_i^{m\prime}, \mathbf{R}_i^*\right)', \tag{7.6}$$

where $\mathbf{Z}_i^0 = (\mathbf{Y}_i^{0\prime}, D_i^0)'$ and $\mathbf{Z}_i^m = (\mathbf{Y}_i^{m\prime}, D_i^m)'$. By grouping the full set into these three components, Njagi et al (2013c) produced a way to represent the information in a form that parallels that for incomplete longitudinal data, with each of the three vectors combining both longitudinal and time-to-event information.

### 7.3.1 The Extended Framework

Similar to Njagi et al (2013c), the following shared random-effects model is considered for this setting:

$$f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^* \mid \mathbf{b}_i) = f(\mathbf{y}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{b}_i) \cdot f(d_i^0 \mid \mathbf{b}_i) \cdot f(d_i^m \mid d_i^0, \mathbf{b}_i) \cdot f(\mathbf{r}_i^* \mid \mathbf{b}_i), \tag{7.7}$$

where $\mathbf{b}_i$ encompasses an elaborate random effects structure, consisting of 31 sets of random effects, i.e., 1 shared between all five components, 5 shared between four components, 10 shared between three components, 10 shared between two components, and 5 specific to a single component. Furthermore, the random effects are assumed independent. The main advantage of obtaining such a modeling framework is that appropriate subsets of random effects can be chosen so that MAR holds (Njagi et al, 2013c).

A few comments are in place. First, it can be noted that model (7.7) is the generic shared random-effects model for this setting under this factorization. Such a general structure implies, that at the time of drop-out, there are processes which may stop, while other processes may get modified. Second, model (7.7) is based on conditional independence assumptions, i.e, given the collection of random effects $\mathbf{b}_i$, the processes $\mathbf{y}_i$, $\mathbf{d}_i$, and $\mathbf{r}_i$ are independent of one another. If all 31 random effects would be present, there still would be a rich association structure present between the various outcomes, which may be simplified by omitting one or more of these components, as will be done to allow for MAR in the next section. For a extended discussion on comments, reference is made to Njagi et al (2013c).

### 7.3.2 Characterization of Missing at Random

The extended model (7.7) allows for a characterization of MAR, in the same spirit as (7.4). To define MAR, either SEM-based or PMM-based factorization of the model can be considered (Njagi et al, 2013c). Under a SEM factorization, the requirement is:

$$f(\mathbf{r}_i^* \mid \mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m) = f(\mathbf{r}_i^* \mid \mathbf{y}_i^0, d_i^0), \tag{7.8}$$

implying that

$$\frac{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m)} = \frac{f(\mathbf{y}_i^0, d_i^0, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, d_i^0)}. \tag{7.9}$$

Under a PMM factorization, the requirement equals:

$$\frac{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, d_i^0, \mathbf{r}_i^*)} = \frac{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m)}{f(\mathbf{y}_i^0, d_i^0)}. \tag{7.10}$$

Using the specific form of (7.7), MAR holds if and only if

$$\frac{\int f(\mathbf{y}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{b}_i) \cdot f(\mathbf{d}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{d}_i^m \mid \mathbf{d}_i^0, \mathbf{b}_i) \cdot f(\mathbf{r}_i^* \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}{\int f(\mathbf{y}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{d}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{r}_i^* \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i} = \frac{\int f(\mathbf{y}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{b}_i) \cdot f(\mathbf{d}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{d}_i^m \mid \mathbf{d}_i^0, \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}{\int f(\mathbf{y}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{d}_i^0 \mid \mathbf{b}_i) \cdot f(\mathbf{b}_i) \cdot d\mathbf{b}_i}. \tag{7.11}$$

### 7.3.3 A Specific Sub-class of the GSPM

Looking for an interesting sub-class of the extended model (7.7), Njagi et al (2013c) proposed the following sub-class:

$$\begin{aligned} f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^* \mid \mathbf{b}_i) = f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i) \cdot f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{m}_i) \cdot f(d_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{l}_i) \\ \cdot f(d_i^m \mid d_i^0, \mathbf{m}_i) \cdot f(\mathbf{r}_i^* \mid \mathbf{g}_i, \mathbf{k}_i, \mathbf{l}_i), \end{aligned} \tag{7.12}$$

where $\mathbf{g}_i$, $\mathbf{h}_i$, $\mathbf{k}_i$, $\mathbf{l}_i$ and $\mathbf{m}_i$ are part of the (earlier described) 31 random effects set. Here, the random effects driving the missing-data components $\mathbf{y}_i^m$ and $d_i^m$ do not appear in any of the other three stochastic components. Furthermore, similar to the extended GSPM (7.7), Njagi et al (2013c) shown that the MAR property is satisfied. Their derivations will be given here.

Denoting $\tilde{\mathbf{b}}_i$ and $\bar{\mathbf{b}}_i$ as the sets of random effects $\{\mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i\}$, and $\{\mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{l}_i\}$, respectively. Under a SEM-based factorization, it follows that

$$f(\mathbf{r}_i^* \mid \mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m) = \frac{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m)} \tag{7.13}$$

$$= \frac{\int_{\tilde{\mathbf{b}}_i} \varphi_1 \cdot \varphi_2 \cdot \varphi_3 \cdot \varphi_4 \cdot \varphi_5 \cdot f(\tilde{\mathbf{b}}_i) \cdot \tilde{\mathbf{b}}_i}{\int_{\mathbf{r}_i^*} \int_{\tilde{\mathbf{b}}_i} \varphi_1 \cdot \varphi_2 \cdot \varphi_3 \cdot \varphi_4 \cdot \varphi_5 \cdot f(\tilde{\mathbf{b}}_i) \cdot d\tilde{\mathbf{b}}_i \cdot d\mathbf{r}_i^*} \tag{7.14}$$

$$= \frac{\int_{\mathbf{m}_i} \varphi_2 \cdot \varphi_4 \cdot f(\mathbf{m}_i) \cdot \mathbf{m}_i}{\int_{\mathbf{m}_i} \varphi_2 \cdot \varphi_4 \cdot f(\mathbf{m}_i) \cdot d\mathbf{b}_i} \cdot \frac{\int_{\bar{\mathbf{b}}_i} \varphi_1 \cdot \varphi_3 \cdot \varphi_5 \cdot f(\bar{\mathbf{b}}_i) \cdot \bar{\mathbf{b}}_i}{\int_{\mathbf{r}_i^*} \int_{\bar{\mathbf{b}}_i} \varphi_1 \cdot \varphi_3 \cdot \varphi_5 \cdot f(\bar{\mathbf{b}}_i) \cdot \bar{\mathbf{b}}_i \cdot d\mathbf{r}_i^*} \tag{7.15}$$

$$= \frac{f(\mathbf{y}_i^0, d_i^0, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, d_i^0)} = f(\mathbf{r}_i^* \mid \mathbf{y}_i^0, d_i^0), \tag{7.16}$$

where

$$\varphi_1 = f(\mathbf{y}_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i) \tag{7.17}$$

$$\varphi_2 = f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{m}_i) \tag{7.18}$$

$$\varphi_3 = f(d_i^0 \mid \mathbf{g}_i, \mathbf{h}_i, \mathbf{l}_i) \tag{7.19}$$

$$\varphi_4 = f(d_i^m \mid d_i^0, \mathbf{m}_i) \tag{7.20}$$

$$\varphi_5 = f(\mathbf{r}_i^* \mid \mathbf{g}_i, \mathbf{k}_i, \mathbf{l}_i) \tag{7.21}$$

From a PMM-based factorization, the following result is pulled forward:

$$f(\mathbf{y}_i^m, d_i^m \mid \mathbf{y}_i^0, d_i^0, \mathbf{r}_i^*) = \frac{f(\mathbf{y}_i^0, \mathbf{y}_i^m, d_i^0, d_i^m, \mathbf{r}_i^*)}{f(\mathbf{y}_i^0, d_i^0, \mathbf{r}_i^*)} \tag{7.22}$$

$$= \frac{\int_{\tilde{\mathbf{b}}_i} \varphi_1 \cdot \varphi_2 \cdot \varphi_3 \cdot \varphi_4 \cdot \varphi_5 \cdot f(\tilde{\mathbf{b}}_i) \cdot \tilde{\mathbf{b}}_i}{\int_{d_i^m} \int_{\mathbf{y}_i^m} \int_{\tilde{\mathbf{b}}_i} \varphi_1 \cdot \varphi_2 \cdot \varphi_3 \cdot \varphi_4 \cdot \varphi_5 \cdot f(\tilde{\mathbf{b}}_i) \cdot d\tilde{\mathbf{b}}_i \cdot d\mathbf{y}_i^m \cdot dd_i^m} \tag{7.23}$$

$$= \frac{\int_{\mathbf{m}_i} \varphi_2 \cdot \varphi_4 \cdot f(\mathbf{m}_i) \cdot \mathbf{m}_i}{\int_{d_i^m} \int_{\mathbf{y}_i^m} \int_{\mathbf{m}_i} \varphi_2 \cdot \varphi_4 \cdot f(\mathbf{m}_i) \cdot d\mathbf{y}_i^m \cdot dd_i^m} \tag{7.24}$$

$$\cdot \frac{\int_{\bar{\mathbf{b}}_i} \varphi_1 \cdot \varphi_3 \cdot \varphi_5 \cdot f(\bar{\mathbf{b}}_i) \cdot \bar{\mathbf{b}}_i}{\int_{\bar{\mathbf{b}}_i} \varphi_1 \cdot \varphi_3 \cdot \varphi_5 \cdot f(\bar{\mathbf{b}}_i) \cdot d\bar{\mathbf{b}}_i} \tag{7.25}$$

$$= f(\mathbf{y}_i^m, d_i^m \mid \mathbf{y}_i^0, d_i^0), \tag{7.26}$$

with $\varphi_1, \ldots, \varphi_5$ formulated in (7.17) − (7.21). Concluding from these derivations, a sufficient condition for the extended model to satisfy MAR is that the random effects influencing the observed measurements and/or the coarsening mechanism do not influence the missing measurements, given the observed measurements, i.e., equivalent to the condition that all information about the missing measurements stems from the observed measurements and covariates only.

### 7.3.4 An MAR Counterpart to an Extended Shared- parameter Joint Model for Longitudinal and Time-to-event Data

From the developments of the previous Sections 7.3.2 and 7.3.3, a door opens to the construction of an MAR counterpart for any member of the extended model (7.7), with exactly the same fit to the observed data, by integrating over the distribution of the missing components given the observed ones (Molenberghs et al, 2008). In other words, to obtain an MAR counterpart for any member of the extended model (7.7), with the same fit to the observed data, distributions $f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{b}_i)$ and $f(d_i^m \mid d_i^0, \mathbf{b}_i)$ in (7.7) need to be changed by

$$\begin{aligned} h(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{m}_i^*) &= \int_{\mathbf{b}_i^*} f(\mathbf{y}_i^m \mid \mathbf{y}_i^0, \mathbf{b}_i) \cdot d\mathbf{b}_i^*, \\ h(d_i^m \mid d_i^0, \mathbf{m}_i^*) &= \int_{\mathbf{b}_i^*} f(d_i^m \mid d_i^0, \mathbf{b}_i) \cdot d\mathbf{b}_i^*, \end{aligned} \tag{7.27}$$

respectively. Integration over $\mathbf{b}_i^*$ equals to the integration over all random effects in the full set $\mathbf{b}_i$, except possibly those that are specific to either $\mathbf{y}_i^m$, or $d_i^m$, or both.

The non-uniqueness does not stem from the model described, but rather from the fact that the model specifies more than what is available in the data (Molenberghs et al, 2012). This implies the need for care when analyzing and interpreting results from models with missing and censored observations, random-effects models, factor-analytic models, etc. Therefore, apart from goodness-of-fit tests, which describe the model fit of the observed data, sensitivity analyses is advisable, i.e., a methodology that studies how assumptions

about unobservables, given the observables, influence the inferences drawn. For example, Kenward, Goetghebeur and Molenberghs (2001) examined sensitivity analysis for incomplete categorical data. In particular, sensitivity analysis for a psychiatric study has been studied, where the study, i.e., a multicentre study, involves 315 patients that were treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonine in the brain (Molenberghs and Lesaffre, 1994). Creemers, et al (2010), on the other hand, proposed a method for sensitivity analysis within the GSPM framework. Specifically, they applied it for a chosen modeling scenario in the toenail data (De Backer et al, 1996), where the methodology used consisted of 5 specific steps. Njagi et al (2013c).

To conclude this chapter, its worthwhile to mention that Njagi et al (2013) additionally examined their research within the extended framework (7.7) by exploring a narrow definition of a joint model, and examining its main limitation, i.e., it defies an MAR characterization. A full discussion on this conclusion can be found in Njagi et al (2013)!

# Chapter 8

# Goodness-of-Fit Test for the Random-Effects Distribution in the Combined Model

In mixed models, consisting of both random and fixed effects, researchers often rely on the assumption that the random effects are normally distributed. Misspecifying the random-effects distribution has modest consequences on maximum likelihood estimators, especially on generalized linear mixed models. For linear mixed models (Section 3.3.1), Verbeke and Lesaffre (1997) showed that the estimators of fixed effects and variance components with normality assumption are consistent and asymptotically normally distributed, even when the true random effects do not follow a normal distribution. Their asymptotic covariance matrix, on the other hand, is biased. Additionally, McCulloch and Neuhaus (2011a & 2011b) and Verbeke and Lesaffre (1996) pointed out that there can be serious consequences on the EB estimation of the random effects. For generalized linear mixed models, Heagerty and Kurland (2001) and Litière, et al (2008) illustrated that the maximum likelihood estimators are inconsistent when the distribution of the random-effects is misspecified. Moreover, increasing the number of random effects expands this problem even more. A quick overview and examples of the consequences of misspecifying the random-effects distribution can be found in Grilli and Rampichini (2014) and Agresti et al (2004).

In the last few years, a lot of research has been devoted to checking distributional assumptions about the random-effects. For example, Waagepetersen (2006) introduced a simulation-based test, by generating random effects while conditioning on the observations, while Tchetgen and Coull (2006) generated a diagnostic test by comparing marginal and conditional maximum likelihood estimators of a subset of fixed effects in the model. Both were created to investigate the suitability of the choice of the random-effects distribution. Even though a feasible power was found with count data for the former one, very large cluster and sample sizes were needed for binary data to produce similar results. The latter one was restricted to those settings where at least one within-cluster covariate is available. Furthermore, several tests has been derived for model misspecification in mixed models. Claeskens and Hart (2009) first explored several formal diagnostics that test the normality assumption of the random effects and/or errors. Adapted traditional tests for normality such as the Pearson $\chi^2$-type tests for mixed models (Jiang, 2001) were mentioned in detail, alongside with some graphical diagnostics for mixed models (Calvin

and Sedransk, 1991 & Lange and Ryan, 1989). Secondly, nonparametric tests based on the order selection concept of Eubank and Hart (1992) were proposed to detect virtually any alternative to normality, where the nonparametric estimation method that is used to construct the test provides an estimator of the alternative distribution (in case of rejection of the null hypothesis). Additionally, White (1982) developed an Information Matrix Test (IMT) for model misspecification. While third-order partial derivatives of the likelihood function are required in this test, this can become a sufficient problem when dealing with complicated likelihood functions, like in generalized linear mixed models. The use of numerical approximations may be needed, which cumbersome the implementation techniques in many statistical software packages. To overcome these problems, Alonso, Litière and Molenberghs (2010) proposed two alternative diagnostic tools, i.e., the so-called Sandwich Estimator Test (SET) and Modified Information Matrix Test (MIMT), to detect misspecification in generalized linear mixed models. These tools use the ideas of the IMT of White (1982), where no third-order partial derivatives of the likelihood are needed. However, they pointed out that the tests also exhibited inflated type I error rates when the sample size was small or moderate. A parametric bootstrap version of the tests seems to overcome this problem, but still require further research.

Alternatively to the testing procedures for checking random-effects distributional assumptions, several efforts have been made in relaxing the parametric assumptions about the random-effects distribution. Zhang and Davidian (2001) used the so-called semi-nonparametric representation of a density function as studied by Gallant and Nychka (1987). To arrive at a proper density function, the estimator appears as a Hermite series, where the normal density function is multiplied by the square of a polynomial and suitably normalized. Moreover, Verbeke and Lesaffre (1996) used a finite mixture of normal density functions to approach the actual random-effects distribution in linear mixed models. Recently, Ghidey, Lesaffre and Verbeke (2010) compared these two approaches, together with two additional approaches, i.e., the smoothing by roughening approach of Shen and Louis (1999) and a flexible approach of Ghidey, Lesaffre and Eilers (2004), via an extensive simulation study.

In this chapter, attention will be given to the gradient function of Verbeke and Molenberghs (2013). This approach serves as a graphical exploratory diagnostic tool to assess misspecification of the random effects distribution, and is applicable to a wide range of mixed models (LMM, GLMM, non-linear mixed models), with univariate as well as multivariate random effects, as long as the conditional distribution for the outcome given the random effects has been correctly specified. Because the tool only requires maximum likelihood estimates of the current model and corresponding marginal likelihood function, easy implementation is present. Moreover, the gradient function is plotted alongside with confidence bands, pointing out intervals of values of the random effects for which the distribution is locally misspecified. Additionally, it indicates how a parametric model can be improved in case of misspecification. Unlike the other mentioned testing procedures for the random-effects distributional assumptions in mixed models, this tool is informal in such a way that it should not be interpreted as a formal testing procedure. Efendi, Drikvandi, Verbeke and Molenberghs (2014) used this tool to develop a simple diagnostic test for the random-effects distribution in mixed models. Specifically, the gradient function will serve as basis for the construction of the proposed formal test.

While Section 8.2 is devoted to the gradient function of Verbeke and Molenberghs (2013), Section 8.3 is used to discuss the formal testing procedure of Efendi, Drikvandi, Verbeke and Molenberghs (2014). As reminder and starting point to this testing paradigm, a general review is given on mixed models (Section 8.1).

## 8.1 Review of the General Mixed Model

Using the same terminology as before (Section 3.3), let $Y_{ij}$ denotes the $j$th measurement for cluster (subject) $i$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and $\mathbf{Y}_i$ represents the $n_i$-dimensional vector of all measurements available for cluster (subject) $i$. In case of time-to-event data, $Y_{ij}$ indicates the $j$th time-to-event outcome for cluster (subject) $i$. In mixed models, similar to Section 3.3, it is assumed that, conditionally on ($q$ x 1)-dimensional vector of random effects $\mathbf{b}_i$, outcome $\mathbf{Y}_i$ follows a particular pre-specified distribution $F_i$, possibly depending on covariates ($\boldsymbol{\xi}$ in previous discussed chapters) and parameterized through a ($p$ x 1)-dimensional vector $\boldsymbol{\theta}$ of unknown parameters, common to all subjects, i.e., $\mathbf{Y}_i \mid \mathbf{b}_i \sim F_i(\boldsymbol{\theta}, \mathbf{b}_i)$. While previous chapters focused on developing an appropriate conditional model $\mathbf{Y}_i \mid \mathbf{b}_i \sim F_i(\boldsymbol{\theta}, \mathbf{b}_i)$, with relevant choices of fixed parameters $\boldsymbol{\xi}$, unknown parameters $\boldsymbol{\theta}$, random effects $\mathbf{b}_i$ and distribution $F_i$ for time-to-event data, this chapter assumes that the conditional model has been correctly specified. Moreover, random-effects $\mathbf{b}_i$ follows a distribution $G$, denoting the between-unit heterogeneity in the population with respect to the distribution of $\mathbf{Y}_i$.

Suppressing dependency on the vector of unknown parameters $\boldsymbol{\theta}$ in notation, let $f_i(\mathbf{y}_i \mid \mathbf{b}_i)$ specify the density function of $\mathbf{y}_i$, conditional on $\mathbf{b}_i$, and corresponding to distribution $F_i$. Likelihood-based inference, e.g., maximum likelihood (Section 4.2.1), for $\boldsymbol{\theta}$ is usually based on the marginal distribution, where the marginal log-likelihood is used as starting point for the derivations. Assuming independency of the units, the marginal log-likelihood function is formulated by

$$\ell(G) = \sum_{i=1}^{N} \ln \left[ \int_b f_i(\mathbf{y}_i \mid \mathbf{b}_i) \cdot dG(\mathbf{b}_i) \right]. \tag{8.1}$$

In case of the WGN model (Chapter 4), $f_i(\mathbf{y}_i \mid \mathbf{b}_i)$ is replaced by formula (4.9), where numerical integration techniques, e.g., adaptive Gaussian quadrature (Molenberghs and Verbeke, 2005), can be used to approximate the integral expression.

Similar to the normal distribution choice for $G$ in the WGN model (Chapter 4), many researchers often assume a parametric specification of distribution $G$, characterized by a vector $\boldsymbol{\varrho}$ of unknown parameters, while others relax the specification even further to a semi parametric and completely nonparametric, e.g., Mallet, 1986; Butler and Louis, 1992, approach. Chen, Zhang and Davidian (2002) relaxed the assumption of normal mixed distributions by specifying that the distribution of random effects belong to a class of smooth densities. The density is thereby approximate by the seminonparametric (SNP) approach of Gallant and Nychka (1987), allowing it to be skewed, multi-modal, fat- or thin-tailed relative to the normal distribution. Moreover, they proposed a Monte Carlo expectation-maximization (EM) algorithm using a rejection sampling scheme to estimate the fixed parameters of the linear predictor, variance components and the SNP density.

To check whether the fitted model with marginal log-likelihood $\ell(\widehat{G})$ adequately fits the data, or whether an alternative mixing distribution, say $H$, could yield a marginal log-likelihood $\ell(\widehat{H})$ substantially larger than $\ell(\widehat{G})$, Verbeke and Molenberghs (2013) proposed the gradient function to graphically check whether $\ell(\widehat{G})$ can be increased substantially by replacing $G$ by $H$, i.e., $\ell(H) > \ell(\widehat{G})$, indicating that the model has been misspecified. A detailed overview of this framework is given in Section 8.2.

## 8.2 The Gradient Function

To obtain a complete overview of the developed terminology behind the gradient function, Verbeke and Molenberghs (2013) started their discussion by defining the directional derivative of the log-likelihood $\ell(\cdot)$ at $G$ into the direction $H$. Due to mathematical properties, this approach will be adopted here.

For two distribution functions $G$ and $H$, the directional derivative of the log-likelihood $\ell(\cdot)$ at $G$ into the direction $H$ is defined as

$$\Phi(G,H) = \lim_{\alpha \gtrdot 0} \frac{\ell[(1-\alpha)\cdot G + \alpha \cdot H] - \ell(G)}{\alpha} = \left.\frac{\partial \ell[(1-\alpha)\cdot G + \alpha \cdot H]}{\partial \alpha}\right|_{\alpha=0}. \tag{8.2}$$

In other words, formula (8.2) denotes the change that is achieved in the log-likelihood by replacing the distribution $G$ by the mixture distribution $(1-\alpha)\cdot G + \alpha \cdot H$ for an infinitesimal weight $\alpha$ assigned to the distribution $H$.

A few comments are in place. First, if $\Phi(\widehat{G}, H) \leq 0$ hold for all $H$, no better mixing distribution than the conducted parametric fit $\widehat{G}$ can be found. Second, for any two distribution functions $G$ and $H$, Verbeke and Molenberghs (2013) showed that the directional derivative of the log-likelihood $\ell(\cdot)$ at $G$ into the direction $H$ is proportional to the gradient function $\Delta(G, \mathbf{b})$, which they defined as

$$\Delta(G, \mathbf{b}) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{f_i(\mathbf{y}_i \mid \mathbf{b}_i)}{f_i(\mathbf{y}_i \mid G)} \tag{8.3}$$

and can be interpreted as an average of likelihood ratios, where each ratio measures how much more likely $\mathbf{y}_i$ is to be observed for unit $i$ if the corresponding random effect $\mathbf{b}_i$ equals $\mathbf{b}$ rather than it being sampled from G. Their explicit derivation can be found in Appendix $E$.1. More specifically, Verbeke and Molenberghs (2013) denoted that

$$\Delta(G, \mathbf{b}) = \frac{1}{N} \cdot \Phi(G, H_{\mathbf{b}}) + 1, \tag{8.4}$$

where $H_{\mathbf{b}}$ represents the discrete distribution function with all probability mass at $\mathbf{b}$. Therefore, $\Delta(G, \mathbf{b})$ can be seen as the score statistic for comparing the 'null model' $\ell(G)$ to an 'alternative model' that would assign additional weight to the support point $\mathbf{b}$. Third, when $\Phi(\widehat{G}, H) \leq 0$, a number of properties about the gradient function $\Delta(G, \mathbf{b})$ of $\widehat{G}$ hold, i.e.,

Property 1. $\boxed{\Phi(\widehat{G}, H) \leq 0 \Leftrightarrow \Delta(\widehat{G}, \mathbf{b}) \leq 1 \text{ for all possible } \mathbf{b}}$

*Proof.*

- $\Rightarrow$ Trivial, since $\Phi(\widehat{G}, H) \leq 0$ also holds for $H = H_{\mathbf{b}}$, i.e., the discrete distribution function with all probability mass at $\mathbf{b}$.
- $\Leftarrow$ Trivial, due to derivation $(E.1)$.

■

Property 2. $\boxed{\Delta(\widehat{G}, \mathbf{b}) = 1 \text{ in all support points } \mathbf{b} \text{ of } \widehat{G}}$

*Proof.* Since $\Phi(\widehat{G}, \widehat{G}) = 0$, it follows that $\int_{\mathbf{b}} \Delta(\widehat{G}, \mathbf{b}) \cdot d\widehat{G}(\mathbf{b}) = 1$. Due to property 1, $\Delta(\widehat{G}, \mathbf{b}) = 1$ in all support points $\mathbf{b}$ of $\widehat{G}$ simply implies. ■

Property 3. $\boxed{\begin{array}{c}\Delta(\widehat{G}, \mathbf{b}) \text{ can only have (local) maxima in the region } I \text{ of } \mathbb{R}^q, \\ \text{equal to the Cartesian product of the intervals } [b_{jmin}, b_{jmax}]\end{array}}$

*Proof.* If all $f_i(\mathbf{y}_i \mid \mathbf{b})$, as functions of $\mathbf{b} = (b_1, \ldots, b_q)'$, have unique modes $\widehat{\mathbf{b}}_i = (\widehat{b}_{i1}, \cdot, \widehat{b}_{iq})'$, then let $[b_{jmin}, b_{jmax}]$ be an interval containing all these mode components $\widehat{b}_{ij}$, $j = 1, \ldots, q$. All $f_i(\mathbf{y}_i \mid \mathbf{b})$ are then monotone increasing functions of $b_j$, whenever $b_j \leq b_{jmin}$ and monotone decreasing whenever $b_j \geq b_{jmax}$, hence the same holds for $\Delta(\widehat{G}, \mathbf{b})$. This implies that $\Delta(\widehat{G}, \mathbf{b})$ can only have (local) maxima in the region $I$ of $\mathbb{R}^q$, equal to the Cartesian product of the intervals $[b_{jmin}, b_{jmax}]$. Intuitively, this implies that the observed data can only provide information about the support of the mixing distribution within the region $I$. ■

With these three properties, Verbeke and Molenberghs (2013) concluded that, if no mixing distribution $H$ provides a better fit than the conducted parametric fit $\widehat{G}$, the gradient function of $\widehat{G}$, i.e., $\Delta(\widehat{G}, \mathbf{b})$, should never exceed 1, and should be exactly equal to 1 in all support points of $\widehat{G}$ that are all in the region $I$. Thus, the fit of a specific distribution function $\widehat{G}$ can simply be checked graphically by inspecting its gradient function $\Delta(\widehat{G}, \mathbf{b})$. If it does not exceed 1, and if it reaches 1 in its support points, then no other mixing distribution $H$ can provide a better fit for the data.

While most frequently used parametric mixed and combined models assume a (multivariate) normal distribution for $G$, a finite sample cannot support the fitted normal distribution $\widehat{G}$ as the best fitting model, because values on the whole real space $\mathbb{R}^q$ are taken, i.e. outside the region $I$. Even though this seems an problematic issue to the developed framework, the gradient function can still be used to check whether the fitted distribution $\widehat{G}$ provides an adequate fit, within the region $I$ (where the data support probability mass for the mixing distribution). If no other mixing distribution $H$ can yield a substantially better fit than the fitted normal $\widehat{G}$, the gradient function $\Delta(\widehat{G}, \mathbf{b})$ should be close to

1 within the region $I$. To distinguish true deviations from 1, from random variability, Verbeke and Molenberghs (2013) suggested to use a pointwise confidence band around $\Delta(\widehat{G}, \mathbf{b})$, because its asymptotic distribution is normal distributed. Hence, due to the central limit theorem (Brown, 2011) with variance estimated by the sample variance of the likelihood ratio contributions $\frac{f_i(\mathbf{y}_i|\mathbf{b}_i)}{f_i(\mathbf{y}_i|G)}$, pointwise confidence limits for $\Delta(\widehat{G}, \mathbf{b})$ can easily be obtained.

At last, to conclude the discussion on the gradient function framework for mixed and combined models, attention is given on an additional powerful aspect of the gradient function. Even when the gradient function obvious exceeds 1, Verbeke and Molenberghs (2013) mentioned that the shape of the gradient function gives some indication of how the distribution $G$ can be adapted to provide a better fit. An increase in log-likelihood can be achieved by replacing the mixing distribution $G$ by $H$, chosen such that $\Phi(G, H) > 0$. It directly follows from derivation $(E.1)$ that $H$ should have considerable support in areas where the gradient function is large (i.e. larger than 1) and little support in areas where the gradient function is small (i.e. smaller than 1). Hence, a model with a gradient function exceeding 1 can be improved by moving probability mass from areas where the gradient function is small to areas where the gradient function is large.

In the next section, the formal testing procedure of Efendi, Drikvandi, Verbeke and Molenberghs (2014) will be discussed. This test is based on the gradient function, i.e., formula (8.3), and can be seen a simple applicable diagnostic test for the random-effects distribution in mixed and combined models, e.g., the WGN model.

## 8.3 The Testing Procedure

While Verbeke and Molenberghs (2013) proposed a simple, powerful graphical tool to check the impact of assumptions about the random-effects distribution in mixed and combined models on inferences, Efendi, Drikvandi, Verbeke and Molenberghs (2014) used this approach to develop a simple diagnostic test for the random-effects distribution in mixed models, where inference is conducted through the bootstrap. This section will be devoted to this approach.

Let $\{b_k, k = 1, \ldots, K\}$ be a sufficiently fine grid in region $I$. Efendi, Drikvandi, Verbeke and Molenberghs (2014) defined the test-statistic by

$$T = \frac{1}{K} \cdot \sum_{k=1}^{K} \left| \widehat{\Delta}(\widehat{G}, b_k) - 1 \right|, \tag{8.5}$$

where $\widehat{\Delta}$ explicitly acknowledges the fact that the unknown parameters $\boldsymbol{\theta}$ in $f_i(\mathbf{y}_i \mid \mathbf{b}_i)$ have been replaced by their estimators $\widehat{\boldsymbol{\theta}}$. From formula (8.5), it can easily be concluded that test-statistic $T$ quantifies the deviation of gradient function $\Delta(\widehat{G}, \mathbf{b})$ from 1, within the interval $I$. The null-distribution of $T$, which is needed to formally test whether the assumed mixing distribution $G$ is appropriate, can be obtained using parametric bootstrap. The following steps are then required in order to perform the bootstrap test:

1. Based on the observed data, fit the mixed model under consideration, with a particular assumption for the mixing distribution $G$, i.e. maximize $\ell(G)$ with respect to the vector $\boldsymbol{\omega}' = (\boldsymbol{\theta}', \boldsymbol{\varrho}')$ of unknown parameters which completely characterizes the marginal density $f_i(\mathbf{y}_i \mid G)$.

2. Construct the gradient function and compute the resulting observed value $T_a$ for the test-statistic $T$.

3. For $s = 1, \ldots, S$, repeat the following steps:

   (a) Sample a new vector $\boldsymbol{\omega}^s$ of parameter values from a multivariate normal distribution with mean $\widehat{\boldsymbol{\omega}}$ and covariance matrix equal to the inverse Fisher information matrix for the fitted model.

   (b) Sample random effects $\mathbf{b}_i^s$, $i = 1, \ldots, N$, from $G$ in which $\boldsymbol{\omega}$ has been replaced by $\boldsymbol{\omega}^s$.

   (c) Sample new observations $\mathbf{Y}_i^s$, $i = 1, \ldots, N$, from $f_i(\mathbf{y}_i \mid G)$ in which $\boldsymbol{\theta}$ has been replaced by $\boldsymbol{\theta}^s$. Note that the data set should have the same structure as the original data set (covariates, number of measurements, etc.)

   (d) Fit the mixed model under consideration based on the sampled data $\mathbf{Y}_i^s$, $i = 1, \ldots, N$.

   (e) Construct the gradient function and compute the resulting observed value $T^s$ for the test-statistic $T$.

4. Calculate the $p$ value as the proportion of values $T^s$ exceeding $T_a$.

Since the construction of interval $I$ depends on the observations in the bootstrap procedure, the interval $I$ changes with each bootstrap sample. Particularly, the interval is determined from knowing the minimum and maximum of the unique modes of all $f_i(y_i \mid b)$ as functions of $b$. The unique modes are calculated through maximizing each $f_i(y_i \mid b)$ (model fitting by subject/cluster) with parameter estimates from maximizing $f(\mathbf{y} \mid b)$ set as offsets except the one related to $b$. Moreover, Efendi, Drikvandi, Verbeke and Molenberghs (2014) evaluated the operating characteristics of the test in a simulation study.

# Chapter 9

# Local Influence Diagnostics for the Combined Model

Commonly, after formulating and fitting a model in statistics, an assessment of model fit and a diagnostic analysis is advisable. Whereas a goodness-of-fit test for the random-effects distribution in mixed and combined models has been discussed in Chapter 8, this chapter focuses on the detection of influential subjects, i.e., a methodology of sensitivity analysis for assessing the influence of small perturbations in a general statistical model. In particularly, attention is given on the location of influential subjects WGN model (Chapter 4).

In the past, several approaches have been discussed for detecting influential observations for (generalized) linear models. For example, for linear regression models, an important approach for identifying influential observations based on case deletion was proposed by Cook (1977a, 1977b & 1979), by defining the so-called Cook's distance. This quantity measures the effect of removing one observation on a parameter estimate or a fitted value. If an observation produces a significant difference in the analysis, the observation is pronounced as an influential point in the data. Nowadays, the Cook's distance measurement has become a popular tool to detect influential observations in linear models by its inclusion in popular statistical software tools such as SAS and R. In the context of missing data, Zhu and Lee (2001) developed a methodology to assess local influence in a minor perturbation of a statistical model with incomplete data, by utilizing Cook's approach to the conditional expectation of the complete-data log-likelihood function in the EM algorithm. It has the potential to assess a variety of complicated models that cannot be handled by existing methods. For elliptical linear models, e.g., the normal, Student t-, Cauchy and logistic distributions, local influence analysis in the univariate case has been made by Galea et al. (1997) and Liu (2000). Liu (2001) introduced a general framework with the concepts of the observed information matrix and the so-called Delta matrix so that the local influence method becomes applicable to elliptical linear regression models in the multivariate case. Delta matrices under the perturbed models of perturbations in case-weights, explanatory variables and response variables are derived, respectively. A general discussion on elliptical models can be found in Frahm (2004) and Lemonte and Patriota (2011).

Extending local influence approaches to mixed and combined models is far from straightforward. For example, linear mixed models, unlike linear models, generally do not allow

for closed-form parameter estimators. Along with other reasons, Lesaffre and Verbeke (1998) therefore proposed a case-weight perturbation scheme for LMM. This scheme investigates how much the parameter estimates are affected by changes in the weights of the log-likelihood contributions of special observations. Several useful characteristics are present in their approach, and its worthwhile to express them. First, influence in fixed-effects parameters are distinguished from that in variance components. Second, for each of these parameter subsets, influence is decomposed in interpretable components. Third, once the model is fitted, the influence diagnostics are computationally inexpensive. For GLMM, less attention has been given on the local influence detection of observations. One of the main complications is that the (log-)likelihood function does not admit a closed form. Hence, their derivations were numerical in nature, which makes it less evident to derive meaningful influence components.

Rakhmawati, Molenberghs, Verbeke and Faes (2014) extended local influence for the GLMM in several ways, by constructing a framework that allows overdispersion in GLMM, i.e., the combined model, and is applicable to binary, count, and time-to-event outcomes. Moreover, three approaches have been suggested, i.e., (1) purely numerical derivations, (2) using a closed-form expression of the marginal likelihood function and (3) using an integral representation of this likelihood. In particularly, this thesis places emphasis on local influence paradigm of the Weibull-normal (WN) model, i.e., the defined Weibull-based GLMM of Section (4.1). Applying and discussing the framework to the WGN model can be done as well.

The chapter is organized as follows. Section 9.1 reviews the essence of local-influence theory. The LMM case of Lesaffre and Verbeke (1998) is described in Section 9.2.1, with the additional proof of Rakhmawati, Molenberghs, Verbeke and Faes (2014) that the integral form of the log-likelihood leads to exactly the same expressions. At last, the local influence paradigm of Rakhmawati, Molenberghs, Verbeke and Faes (2014) is conducted to the WN model (Section 9.2.2).

## 9.1 General Theory of Local Influence

### 9.1.1 Standard Approach

Local influence has become an important step in the analysis of a dataset. While Cook (1986) popularized this paradigm in statistics, lot of research has been devoted ever since. The origin of local influence discussion lies in Cook's (1977) paper, and will serve as starting point.

Cook's (1977) paper emphasizes case-deletion diagnostics, i.e., a popular way to assess the individual impact of cases on the estimation process, for all kind of models. While these diagnostics can be placed in a global influence analysis, where the effect of an observation is assessed by completely removing it, local influence analysis have been discussed later on (Cook, 1986). This approach gives a weight $\omega_i$ to each observation and measures the effect on the parameter estimation by perturbing these weights around, say, $\omega_i = 1$. Lesaffre and Verbeke (1998) pointed out that the choose of weight equally to 0 or 1 corresponds to the global case-deletion approach. Using Cook's (1986) paper as reference, Lesaffre and

Verbeke (1998) introduced influence assessment for the linear mixed model. Even while technical requirements are needed for local influence, it leads to easy and fast calculations and even to interpretable components of influence (in many cases). A review of several diagnostic procedures for the linear mixed model is given in Mun and Lindstrom (2013).

Rakhmawati, Molenberghs, Verbeke and Faes (2014) mainly focused on the paper of Ouwens, Tan, and Berger (2001), where local influence was explored in the Poisson-normal model. Rakhmawati et al (2014) extended their framework in three ways (see above), and developed it for the combined model framework in three particular cases: binary, count, and time-to-event data.

Using the same terminology as before (Section 8.1), let $\ell(\boldsymbol{\theta})$ denote the log-likelihood for the generalized linear mixed or combined model, and is represented by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}), \tag{9.1}$$

where $\ell_i(\boldsymbol{\theta})$ represents the contribution of the $i$th observation to the log-likelihood function with vector of unknown parameters $\boldsymbol{\theta}$. Cook (1986) extended this terminology by defining $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ as the perturbed version of $\ell(\boldsymbol{\theta})$, depending on an $N$-dimensional weight vector $\boldsymbol{\omega}$ that belongs to an open subset $\Omega$ of $\mathbb{R}^N$. Mathematically,

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \sum_{i=1}^{N} \omega_i \cdot \ell_i(\boldsymbol{\theta}). \tag{9.2}$$

By comparing formula (9.2) and (9.1), there exists a weight vector $\boldsymbol{\omega}_0$ such that $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \ell(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$, i.e., $\boldsymbol{\omega}_0 = (1, 1, \ldots, 1)'$. Furthermore, let $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{\theta}}$ be the maximum likelihood estimators for $\boldsymbol{\theta}_{\boldsymbol{\omega}}$ and $\boldsymbol{\theta}$, respectively. To assess the influence of varying $\omega$ throughout $\Omega$, Cook (1986) suggested to measure the distance between $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{\theta}}$ by the so-called likelihood displacement, i.e.,

$$\text{LD}(\boldsymbol{\omega}) = 2 \cdot \left( \ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) \right). \tag{9.3}$$

Here, a few comments are in place. First, Rakhmawati, Molenberghs, Verbeke and Faes (2014) pointed out that $\text{LD}(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\theta})$ is strongly curved at $\boldsymbol{\theta}$. Secondly, Cook (1986) mentioned that information on the influence of case-weight perturbations are brought out by a graph of $\text{LD}(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$, i.e., the geometric surface formed by the values of the $(N + 1)$ x 1 dimensional matrix

$$\boldsymbol{\chi}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ \text{LD}(\boldsymbol{\omega}) \end{pmatrix} \tag{9.4}$$

as $\boldsymbol{\omega}$ varies throughout the open subset $\Omega$ of $\mathbb{R}^N$. In mathematics, more specifically differential geometry, surfaces of this kind are often called Monge patch's (Millman and Parker, 1977), and have been widely used as basic entity in many practical (data science)

applications, e.g., for facial feature tracking with 3D TOF cameras (Haker et al, 2007). For conventional reasons (Cook, 1986), $\boldsymbol{\chi}(\boldsymbol{\omega})$ is referenced as an influence graph.

To characterize the behavior of an influence graph around $\boldsymbol{\omega}_0$, Cook (1986) used geometric normal curvatures that were developed by the use of vector $\boldsymbol{\chi}(\boldsymbol{\omega})$ of formula (9.4). While Cook (1986) also discussed other types of influence graphs, this section only focuses on the first approach, i.e., for $\boldsymbol{\chi}(\boldsymbol{\omega})$ of formula (9.4).

By using the properties that (1) $\mathrm{LD}(\boldsymbol{\omega})$ achieves a local minimum at $\boldsymbol{\omega}_0$, (2) $\partial \mathrm{LD}(\boldsymbol{\omega})/\partial\omega_i = 0$ holds for all $i$, $i = 1, \ldots, N$,

$$(3)\ \left.\frac{\partial \ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})}{\partial \theta_j}\right|_{\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}} = 0, \tag{9.5}$$

for $j = 1, \ldots, p$ and all possible vectors $\boldsymbol{\omega}$ in $\Omega$, (4) applying the chain rule for differentiation twice and (5) differentiating both sides of (9.5) w.r.t. $\boldsymbol{\omega}$ and evaluating at $\boldsymbol{\omega}_0$, Cook (1986) derived that the normal curvature $C_{\mathbf{l}}$ of the lifted line $\boldsymbol{\omega}(x) = \boldsymbol{\omega}_0 + x \cdot \mathbf{l}$ in direction $\mathbf{l}$ equals to

$$C_{\mathbf{l}} = 2 \cdot \left| \mathbf{l}' \cdot \Delta' \cdot \ddot{L}^{-1} \cdot \Delta \cdot \mathbf{l} \right|, \tag{9.6}$$

where $\Delta$ be the $p$ x $N$ matrix with $\boldsymbol{\Delta}_i$, i.e., the $p$-dimensional vector of second-order derivates of $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ w.r.t. $\omega_i$ and all components of $\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, in the $i$th column, $\ddot{L}$ represent the $p$ x $p$ matrix of second derivatives of $\ell(\boldsymbol{\theta})$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, $\mathbf{l}$ denotes a fixed nonzero vector of unit length in $\mathbb{R}^N$ and $x \in \mathbb{R}$.

From formula (9.6), different choices for $\mathbf{l}$ can be made. For example, Rakhmawati, Molenberghs, Verbeke and Faes (2014) mainly focused on subject $i$ only, by choosing $\mathbf{l} = \mathbf{l}_i$, i.e., a zero $N$-dimensional vector with a sole 1 on the $i$th position, and when only a sub-vector $\boldsymbol{\theta}_1$ of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ is of interest. The normal curvature for the former one can simply be expressed by

$$C_i \equiv C_{\mathbf{l}_i} = 2 \cdot \left| \boldsymbol{\Delta}_i' \cdot \ddot{L}^{-1} \cdot \boldsymbol{\Delta}_i \right|. \tag{9.7}$$

For the latter one (Verbeke and Molenberghs, 2000), the normal curvature is formulated by

$$C_{\mathbf{l}}(\boldsymbol{\theta}_1) = C_{\mathbf{l}} + 2 \cdot \mathbf{l}' \cdot \Delta' \cdot \begin{pmatrix} 0 & 0 \\ 0 & \ddot{L}_{22}^{-1} \end{pmatrix} \cdot \Delta \cdot \mathbf{l} \leq C_{\mathbf{l}}. \tag{9.8}$$

When $\ddot{L}_{12}^{-1} = 0$ hold, Rakhmawati, Molenberghs, Verbeke and Faes (2014) pointed out that an influence decomposition is possible:

$$C_{\mathbf{l}} = C_{\mathbf{l}}(\boldsymbol{\theta}_1) + C_{\mathbf{l}}(\boldsymbol{\theta}_2). \tag{9.9}$$

Furthermore, formula (9.9) approximately holds for weakly correlated sub-vectors.

### 9.1.2 Proceeding when faced with a Complicated Likelihood

In Section 9.2.1, attention is given to the local influence framework of Lesaffre and Verbeke (1998). These authors derived local influence based on the explicit expression of the marginalized linear mixed model (see Section 3.3.1). Due to the occurrence of marginal closed expressions for Poisson-Normal, Poisson-Gamma-Normal, Probit-Normal and WN models, Rakhmawati, Molenberghs, Verbeke and Faes (2014) extended their approach in two alternative ways: (1) Using integral expression (3.31), combined with the property that integration and derivation can be interchanged under mild regularity conditions, and (2) choosing a fully numerical route, as in Ouwens, Tan, and Berger (2001). In Section 9.2.2, both routes are explored for the WN model.

## 9.2 Local Influence for Generalized Linear mixed Models

### 9.2.1 Local Influence for the Linear Mixed Model

To discuss the local influence paradigm of Rakhmawati, Molenberghs, Verbeke and Faes (2014), the local influence framework of Lesaffre and Verbeke (1998) will be used as starting point. Using the same terminology as before (Section 3.3.1) and to stay in line with these authors, the marginal linear mixed model (3.24) will be studied, with conditional independence assumption $\Sigma_i = \sigma^2 \cdot I_{n_i}$, and $I_{n_i}$ the $n_i$ x $n_i$ identity matrix.

To derive a local influence framework for linear mixed models, Lesaffre and Verbeke (1998) first reformulated expression (9.7) of the normal curvature $C_i$ by

$$C_i = -2 \cdot \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^1_{(i)}\right)' \cdot \ddot{L}_{(i)} \cdot \ddot{L}^{-1} \cdot \ddot{L}_{(i)} \cdot \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^1_{(i)}\right). \tag{9.10}$$

A full explanation of the components can be found in Appendix ($F$.1.1). It is advantageous that $C_i$ admits a closed form (9.7). Assuming $\boldsymbol{\mathcal{R}}_i$, $\mathcal{X}_i$, and $\mathcal{Z}_i$ to be the "standardized" residuals and covariates for the $i$th individual, defined by $\boldsymbol{\mathcal{R}}_i = V_i^{-1/2} \cdot \mathbf{r}_i$, $\mathcal{X}_i = V_i^{-1/2} \cdot X_i$, and $\mathcal{Z}_i = V_i^{-1/2} \cdot Z_i$, respectively, with $\mathbf{r}_i = \mathbf{y}_i - X_i \cdot \widehat{\boldsymbol{\beta}}$ and $V_i = Z_i \cdot D \cdot Z_i' + \sigma^2 \cdot I_{n_i}$. Furthermore, let $\|A\| = \sqrt{\text{tr}(A' \cdot A)}$ be the Frobenius norm of a matrix $A$ (Golub and Van Loan, 1989). With these elements defined, Lesaffre and Verbeke (1998) decomposed $C_i$ into the following five interpretable components:

$$\|\mathcal{X}_i \cdot \mathcal{X}_i'\|, \qquad \|\boldsymbol{\mathcal{R}}_i\|, \qquad \|\mathcal{Z}_i \cdot \mathcal{Z}_i'\|, \qquad \|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i'\|, \qquad \|V_i^{-1}\|. \tag{9.11}$$

A few comments are in place. First, $\|\mathcal{X}_i \cdot \mathcal{X}_i'\|$ measures the length of the standardized covariates in the mean structure and $\|\boldsymbol{\mathcal{R}}_i\|$ represents an overall measure for how well the observed data for the $i$th subject are predicted by the mean structure $X_i \cdot \widehat{\boldsymbol{\beta}}$. Second, a similar meaning is present for the components $\|\mathcal{Z}_i \cdot \mathcal{Z}_i'\|$ and $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i'\|$, respectively, but now for the covariance structure. For example, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i'\|$ equals to zero only if $V_i$ equals $\mathbf{r}_i \cdot \mathbf{r}_i'$, i.e., an estimate for $\text{Var}(\mathbf{y}_i)$, which only assumes the mean to be correctly modeled as $X_i \cdot \boldsymbol{\beta}$. Therefore, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i'\|$ can be understood as a residual, capturing

how well the covariance structure of the data is modeled by $V_i$. Third, $\|V_i^{-1}\|$ becomes large if $V_i$ has small eigenvalues, which indicates that the $i$th subject is assumed to have small variability.

Due to their decomposition of $C_i$, a practical procedure to find an explanation for the influential nature of an individual immediately follows, i.e., when $C_i$ is large, the diagnostics are explored. Such plots are useful to graphically inspect the individuals in view of their influential nature. Lesaffre and Verbeke (1998) therefore suggested to start with an index plot of $C_i$. Particularly, the index plots of the components (9.11) can be investigated. However, a recurrent difficulty with diagnostics is to establish a threshold above which an individual is defined as remarkable. It follows from (9.7) that

$$\sum_{i=1}^{N} C_i = -2 \cdot \text{tr}\left(\ddot{L}^{-1} \cdot \sum_{i=1}^{N} \Delta_i \cdot \Delta_i^{'}\right) \rightarrow 2 \cdot s, \tag{9.12}$$

for $N$ approaching to $\infty$. As for leverage in linear regression, Lesaffre and Verbeke (1998) denoted that one could classify an individual for which $C_i$ is larger than twice the average value (larger than $4 \cdot s/N$, for $N$ large) as being influential. Unlike the leverage situation, however, $2 \cdot s$ is only the approximate sum of the $C_i$. Therefore, non accurate conclusions will be obtained if (1) the model is not correctly specified (such that $\ddot{L}^{-1} \cdot \sum_{i=1}^{N} \Delta_i \cdot \Delta_i^{'}$ does not converge to $I_s$) or (2) if $N$ is too small for the asymptotic results to yield good approximations. To resolve this issue, Lesaffre and Verbeke (1998) proposed to replace $2 \cdot s$ by the actual sum, and called the $i$th subject influential if $C_i$ is larger than the cutoff value $2 \cdot \sum_{i=1}^{N} C_i/N$.

Given decomposition (9.11), Lesaffre and Verbeke (1998) additionally considered subvectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of fixed effects and variance components, respectively, with corresponding influences $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{\alpha})$, respectively. Given that the fixed effects and variance components are asymptotically independent, it follows that

$$C_i \approx C_i(\boldsymbol{\beta}) + C_i(\boldsymbol{\alpha}). \tag{9.13}$$

Lesaffre and Verbeke (1998) showed in this setting that $C_i(\boldsymbol{\beta})$ can be decomposed using only the first two components of (9.11), i.e., $\|\mathcal{X}_i \cdot \mathcal{X}_i^{'}\|$ and $\|\boldsymbol{\mathcal{R}}_i\|$, while the last three components of (9.11), i.e., $\|\mathcal{Z}_i \cdot \mathcal{Z}_i^{'}\|$, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|$ and $\|V_i^{-1}\|$, feature in the decomposition of $C_i(\boldsymbol{\alpha})$. Asymptotically therefore, influence for the fixed effects can be studied by the first two components of (9.11), while the last three components of (9.11) are only needed for the influence of the variance components.

Until now, the standard approach of Lesaffre and Verbeke (1998), based on the marginal likelihood of the LMM, was discussed. Rakhmawati, Molenberghs, Verbeke and Faes (2014) proposed an alternative way, i.e., the so-called integral-based approach, to alleviate complexities with the explicit marginal likelihood expressions. While these authors mainly focused on developments of the Poisson, probit, logit, and Weibull cases, the LMM framework was first used as starting point. Derivations of the integral-based approach for the LMM setting can be found in Appendix ($F$.1.2). As outcome, the same result is obtained. The same interpretable components as in (9.11) ensue.

### 9.2.2 Local Influence for the Weibull-Normal Model

In this section, local influence for the WN model is studied, according to the developed framework of Rakhmawati, Molenberghs, Verbeke and Faes (2014). Both routes, i.e., (1) integral-based and (2) fully numerical approach, will be discussed in detail. An extended overview of the local influence paradigm and corresponding calculations in other generalized linear mixed models, e.g., the Poisson-normal, probit-normal and logit-normal model, can be found in the paper of Rakhmawati, Molenberghs, Verbeke and Faes (2014).

Equivalent to the described WGN model (4.1)–(4.3), the WN model can be formulated by

$$f(\mathbf{y}_i \mid \mathbf{b}_i) = \prod_{j=1}^{n_i} \lambda \cdot \rho \cdot y_{ij}^{\rho-1} \cdot e^{\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i} \cdot e^{-\lambda \cdot y_{ij}^{\rho} \cdot e^{\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i}}, \tag{9.14}$$

$$f(\mathbf{b}_i) = \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i}. \tag{9.15}$$

1. *Integral-based approach*

   Keeping in mind that the joint distribution of the WN model equals ($B.5$), where $\theta_{ij}$ simplifies to 1, and following the same integral-approach of the LMM (Appendix F.1.2), Rakhmawati, Molenberghs, Verbeke and Faes (2014) derived the partial derivatives w.r.t. the fixed effects $\boldsymbol{\xi}$ and variance-covariance matrix $D$ as

$$\frac{\partial l_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\xi}} = \sum_{j=1}^{n_i} \mathbf{x}_{ij} - \lambda \cdot \sum_{j=1}^{n_i} y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \exp(\boldsymbol{\mu}_{ij}), \tag{9.16}$$

$$\frac{\partial l_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot \left[ (D^{-1})_{jk} - (D^{-1} \cdot D^{-1})_{jk} \cdot \mathrm{Var}(\mathbf{b}_i) \right], \tag{9.17}$$

   where $d_{jk}$ is the $(j,k)$ element of $D$, $\boldsymbol{\mu}_{ij} = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i$, $\mathrm{Var}(\mathbf{b}_i) = \sum_{k=1}^{q} \mathrm{Var}(b_{ik})$ and $\delta_{jk} = 1$ if $j = k$ and 0 otherwise.

   From (9.16)–(9.17), interpretable expression can be derived. Using the same terminology as before (Section 9.1.1), Rakhmawati, Molenberghs, Verbeke and Faes (2014) showed that the sum of squares of the contributions for the ith individual, i.e., $\|\boldsymbol{\Delta}_i\|$, equals

$$\begin{aligned} \|\boldsymbol{\Delta}_i\| = & \left( \sum_{j=1} n_i \mathbf{x}_{ij} \right) \cdot \left( \sum_{j=1} n_i \mathbf{x}_{ij} \right)^{'} - 2 \cdot \sum_{j=1} n_i \mathbf{x}_{ij} \cdot \mathbf{Q}_i^{'} + \mathbf{Q}_i \cdot \mathbf{Q}_i^{'} \\ & + \sum_{k,l} \left[ -\frac{1}{2} \cdot (D^{-1})_{kl} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i) \right]^2, \end{aligned} \tag{9.18}$$

   where $\mathbf{Q}_i = \lambda \cdot \sum_{j=1} n_i y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \exp(\boldsymbol{\mu}_{ij})$. They rewrote $C_i$ by the sum of $C_{1i}$ and $C_{2i}$, with

$$
\begin{aligned}
C_{1i} &= 2 \cdot \|\ddot{L}^{-1}\| \cdot \left(\|\mathbf{x}_i\|^2 - 2 \cdot \mathbf{x}_i \cdot \mathbf{Q}_i + \|\mathbf{Q}_i\|^2\right) \cdot \cos(\varphi_i), \\
C_{2i} &= \frac{1}{2} \cdot \|\ddot{L}^{-1}\| \cdot \cos(\varphi_i) \cdot \operatorname{tr}\left[(D^{-1})^2_{kl}\right] - \operatorname{tr}\left[2 \cdot (D^{-1})_{kl} \cdot (D^{-1} \cdot D^{-1})_{kl} \cdot \operatorname{Var}(\mathbf{b}_i)\right] \\
&\quad + \operatorname{tr}\left[(D^{-1} \cdot D^{-1})^2_{kl} \cdot \operatorname{Var}(\mathbf{b}_i)^2\right]. \qquad (9.19)
\end{aligned}
$$

where $\mathbf{x}_i = \sum_{j=1} n_i \mathbf{x}_{ij}$. Note that $C_{1i}$ and $C_{2i}$ are the contributions of the $i$th subject to local influence contributions $C_i$ from $\boldsymbol{\xi}$ and $D$, respectively.

Hence, the interpretable components of $C_i$ for the Weibull normal model can be described using the length of fixed effect ($\|\mathbf{x}_i\|^2$) and the squared of random effect variability ($\operatorname{Var}(\mathbf{b}_i)^2$). Similar results were obtained for the Poisson-normal and logit-normal model (Rakhmawati, Molenberghs, Verbeke and Faes, 2014).

2. *Fully numerical route*

Alternatively, a fully numerically route can be taken. This approach is based on replacing derivatives by appropriately precise finite differences of the first and second order, for the score vector and Hessian matrix, respectively. Since formula (9.7) mainly consists of first- and second-order derivatives of the loglikelihood function, a fully numerically route can be followed for the discussed local influence paradigm. Additionally, easy implementation is present for models with perturbation scheme (9.2) for the log-likelihood, provided that the score and Hessian functions are numerically available. In many statistical software packages, e.g., SAS, such calculations are routinely done in the log-likelihood maximization process. For the score, individual subjects' contributions are needed (see formula (9.7)).

To conclude the discussion on local influence, Rakhmawati, Molenberghs, Verbeke and Faes (2014) pointed out that the use of both routes, i.e., (1) integral-based and (2) fully numerically, are beneficial, due to the computational ease of the latter approach and the explicit calculated fields to understand the influence components more in detail in the first one.

# Chapter 10

# Enriched-data Problems and Essential Non-identifiability

In statistics, two principal ways are present in which statistical models extend beyond the data available. First, the data may be coarsened, i.e., what is actually observed is less detailed than what is planned. Typical causes of coarsening can be attrition, censoring, grouping, or a combination of these. Second, the data may be augmented, i.e., the observed data are hypothetically but conveniently supplemented with structures. These structures can be random effects, latent variables, latent classes, or component membership in mixture distributions. Combining both aspects together is referred as enriched data, and needed to be handled with caution.

Main reasons for modeling enriched data include the incorporation of substantive information, e.g., the need for predictions, advantages in interpretation, and mathematical and computational convenience, where the fitting combines evidence arising from empirical data with non-verifiable model components, i.e., that are purely assumption driven. Therefore, discretion of the potential dangers and pitfalls that follow from this should be present in the analysis. In the past, attention has been given on the missing data and random effects models case (Verbeke and Molenberghs, 2010). Molenberghs, Njeru Njagi, Kenward, and Verbeke (2012) extended this discussion in a much broader framework, encompassing a bigger number of seemingly disparate enriched-data settings.

This chapter focuses on both papers. In Section 10.2, the general results, concerning enriched data structures, are discussed. These results are mainly based on Verbeke and Molenberghs (2010). In particular, attention is given to the fact that the components of the models can be chosen in an effectively infinite number of ways without affecting the fit to the observed data. Moreover, these general results are then applied in two widely used, recognized statistical settings, i.e., random effects models (Section 10.3) and incomplete data (Chapter 7). An extended version of applications can be explored in Molenberghs, Njagi, Kenward, and Verbeke (2012).

## 10.1 Introduction

Statistical models often rely on assumptions that cannot be examined from the data under analysis. Therefore, it is important that the use of these models properly reflects

the implied reliance on external information. Molenberghs, Njagi, Kenward, and Verbeke (2012) provided a good, clear example of the failure to appreciate the nature of such models by the well-known historical developments surrounding factor analysis in so-called general intelligence measurement (Gould, 1981). Even while this thesis does not focuses on factor analysis, the example will be adopted in order to better understand the increased discussion on enriched data.

Factor analysis has a long history in psychology, dating back to the work of Charles Spearman and Karl Pearson in the early 1900's. At that time, psychologists speculated that intelligence could be defined by a single, all-in-one unobservable entity $g$. In Spearman's (1904) paper, Spearman sought to describe the influence of $g$ on examinees test scores on several domains, e.g., mathematics, language, etc. Motivated by the positive correlations among these tests, Spearman used the technique to develop the so-called "two-factor" theory, implying that a set of mental tests represents an underlying general factor ($g$ here), in addition to each test's specific information. Moreover, Spearman named the entity $g$ general intelligence, and declared it further as an attribute, resident in the brain, which he called general energy, alongside with the test-specific information, called $s$-factors, which were identified as specific engines in the brain, which are under the influence of the general energy.

Nowadays, two major schools of thought are present on the nature of intelligence. For the first one, supported by such psychologists as Eysenck, Galton, Jensen, and Spearman, major believe is present that all intelligence comes from one general factor $g$. The proponents of the other school of thought, presented by Gardner, Sternberg, and Thurstone, pronounced that all intelligence cannot be displayed by one single factor, but by more than one general type of intelligence. For example, Thurstone faulted Spearmans methodology of one single factor $g$ by proposing a solution that is a rotation of Spearmans principal-components solution. While the observed data is fitted equally well with the solution of Spearman, difference only arise in aspects of the model that cannot be verified from the data. The value of their respective solutions including their non-verifiable assumptions rests entirely on practical considerations. Molenberghs, Njagi, Kenward, and Verbeke (2012) used this example to show that this phenomenon is very common throughout statistical modeling, and extends across a whole range of common data-analytic approaches. Nevertheless, they used it to indicate the importance of sensitivity in statistical modeling.

Molenberghs, Njagi, Kenward, and Verbeke (2012) distinguish two types of settings. The first one, termed augmented data, applies when the observed data is supplemented with latent or unobserved quantities. Examples include random-effects models, latent class and latent variable models, and finite-mixture models. The second one, referred to as coarsening, was first introduced by Heitjan (Heitjan and Rubin, 1991; Zhang and Heitjan, 2007), and refers to situations where the observed data are coarser than the hypothetically conceived data structures, to which the models of interest apply. Examples include incomplete data and censored survival data. Molenberghs, Njagi, Kenward, and Verbeke (2012) pointed out that there is a formal distinction between the two types. In the coarse-data setting, a part of the data would ideally be observed but is not in practice, e.g., actual survival time after censoring and outcomes after dropout. Augmented data, on the other hand, refers rather to the addition of useful but artificial constructs to the data setting,

e.g., random effects, latent classes, latent variables, factors, and mixture component membership, and can never be observed. They focus on the review of a selected range of each setting, and bring out commonality. Combining both settings together refers to enriched data, and will be treated in a unified way, such that important, common features can be illuminated and scrutinized.

In the parts that follow, two relevant settings will be explored in detail in this thesis, i.e., random effects models and incomplete data. Since a discussion on incomplete data has already been provided in Chapter 7, the random effects models will be briefly discussed here (Section 10.3). Both are discussed in this thesis since a part of the model is totally unidentifiable from the observed data, implying that the identification of such a part can come from assumptions only. Nevertheless, random-effects and missing data are two aspects that form an important part of this thesis. More specifically, and similar to Molenberghs, Njagi, Kenward, and Verbeke (2012), this chapter can be split into the part. First, a setting is provided to show how models in enriched-data environments are identified by a triple framework, i.e., data, design-based assumptions (e.g., randomization), and further unverifiable assumptions. Focus is laid on the model itself and its relationship to the data through likelihood, and not specifically with subsequent inferences and whether a Bayesian or frequentist route is taken. In the Bayesian approach, a part of the model is considered for which the posterior density depends only on the choice of prior density (assuming appropriate independence relationships among components of the prior density), and in the frequentist paradigm that does not affect goodness-of-fit to the observed data. Second, while various forms of this are known in various sub-fields, to variable degrees, emphasis is placed on the great similarity between these fields and settings; appropriate review of a number of selected areas is presented to facilitate study of the common features. This is presented by showing how non-identified parts can be replaced arbitrarily, without altering the fit to the observed data but with potentially non-trivial consequences for inferences and substantive conclusions. It should be clear that this can be dangerous and the user must carefully reflect on the arbitrary components. Molenberghs, Njagi, Kenward, and Verbeke (2012) mentioned that they should be supported by substantive considerations or be made part of a sensitivity analysis, similar to the conclusion of Chapter 7. Therefore, acceptable goodness-of-fit to the observed data cannot be used as the sole justification for the analysis. In the absence of external corroborating knowledge or information, two alternative routes can be followed. First, it can be made clear that the conclusions drawn have meaning only under the external assumptions built into the analysis. For example, a researcher can choose to draw inferences given a set of scientifically plausible but otherwise non-verifiable causal relationships. It is then important not to divorce the data analysis from the assumptions made. Second, an appropriate sensitivity analysis can be conducted to augment the conclusions. By sensitivity analysis, either a study of how unverifiable assumptions affect overall inferences, or an assessment of traceability (Molenaar, 2004, 2008), i.e., how unverifiable assumptions influence predictions for individual subjects. For example, analyses can be conducted under a number of alternative sets of hypothesized structures as well. This then allows the researcher to examine the sensitivity of the inferences concerning the scientific question to varying the underlying assumptions.

## 10.2 General Result about Counterparts in Enriched-data Structures

This section mainly focuses on the developments of Verbeke and Molenberghs (2010), and will be used as basic building block in the process development of the discussion on counterparts in enriched-data structures. Let $\mathbf{Z}_i$ be the data for an independent unit $i$, $i = 1, \ldots, N$, that is augmented with $\mathbf{c}_i$, i.e., any type of enriched-data form. Often encountered examples for vector $\mathbf{c}_i$ are missing data, random effects, or even a combination of both. Here, the column dimension of $\mathbf{Z}_i$ and $\mathbf{c}_i$ (not denoted here) is suppressed from notation for simplicity.

Assume a joint model of the generic form $f(\mathbf{z}_i, \mathbf{c}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi})$, where covariates are suppressed for notational simplicity. Moreover, the parameters are considered to be disjoint, meaning that the parameter space of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ equals the set theoretic product of the individual parameter spaces (Rubin, 1976). Now, consider the following factorizations:

$$f(\mathbf{z}_i, \mathbf{c}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{z}_i \mid \mathbf{c}_i, \boldsymbol{\theta}) \cdot f(\mathbf{c}_i \mid \boldsymbol{\psi}), \tag{10.1}$$

$$= f(\mathbf{z}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi}) \cdot f(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}). \tag{10.2}$$

By using the same terminology as before (Chapter 3, 4 & 5), names can be given to every factor of (10.1) and (10.2). The left part of equation (10.1) and (10.2) equals to the joint model. For the right hand side of (10.1), a split is made. The first part, i.e., $f(\mathbf{z}_i \mid \mathbf{c}_i, \boldsymbol{\theta})$, can be labeled as the hierarchical model, while the second part, i.e., $f(\mathbf{c}_i \mid \boldsymbol{\psi})$, refers to the prior density for the enriched data. An analogue categorization can be given to the right hand side of (10.2). The first factor in (10.2), i.e., $f(\mathbf{z}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi})$, may be termed the marginal model, whereas the second one, i.e., $f(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\theta}, \boldsymbol{\psi})$, refers to the posterior density of the enriched data.

From (10.1) – (10.2) and the mixed-model setting (Section 8.1), an obvious link is present. The link with incomplete data follows by setting $\mathbf{c}_i \equiv \mathbf{y}_i^m$ and $\mathbf{z}_i = (\mathbf{y}_i^0, \mathbf{r}_i)$. These considerations immediately establish the following theorem:

**Theorem 10.2.1** (A Family of Counterparts to a Given Model for Enriched Data)**.** *Let data $\mathbf{z}_i$ be enriched with $\mathbf{c}_i$. Then, any model* (10.1) *formulated for and fitted to such data, can be replaced by an infinite family of models, all retaining the fit to the observed data as achieved by the original model. This is done by preserving the marginal model $f(\mathbf{z}_i \mid \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$ and replacing the posterior density $f(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\theta}, \boldsymbol{\psi})$ by an arbitrary conditional density*

$$f(\mathbf{d}_i \mid \mathbf{z}_i, \boldsymbol{\gamma}). \tag{10.3}$$

Some comments are in place here. First, the vector $\mathbf{d}_i$ indicates that there not need to be any connection between the original and substituted enriched data, contrary to $\mathbf{c}_i$. Additionally, the (new) density (10.3) can be parameterized by a completely new parameter $\boldsymbol{\gamma}$.

## 10.3 Random Effects Models

To stay in line with this thesis' topic, a discussion around random effects models is made. More particularly, and similar to Molenberghs, Njagi, Kenward, and Verbeke (2012) and Verbeke and Molenberghs (2010), the standard linear mixed model of Section 3.3.1 will be considered (Section 10.3.1). Verbeke and Molenberghs (2010) extended their discussion on random effects in LMM by exploring a special but enlightening case of exchangeable, compound symmetry data, in the sense that all members of a cluster have the same mean $\mu_i$ and the variance-covariance matrix is of a compound-symmetry structure, i.e., $V_i = \sigma^2 \cdot I_{n_i} + d \cdot J_{n_i}$, where $I_{n_i}$ is an $n_i$-dimensional identity matrix and $J_{n_i}$ is an $n_i$ x $n_i$ matrix consisting of 1's. This setting is referred as the 'exchangeable' one (Section 10.3.2), equivalent to Verbeke and Molenberghs (2010).

### 10.3.1 The Standard Linear Mixed Models

In order to apply Theorem 10.2.1 to the LMM setting (Section 3.3.1), formulation (3.22) - (3.24) will be used to first express all components featuring in (10.1) - (10.2), i.e., hierarchical model, prior density, marginal model and posterior density, for the LMM framework. Secondly, the posterior density for the random effects, which is often chosen normal, is replaced by two versions of the exponential density.

1. *Components featuring in* (10.1) - (10.2)

   Using formulation (3.22) - (3.24) of the LMM, the fully hierarchically specified linear mixed effects model is formulated by (Verbeke and Molenberghs, 2000)

   $$\mathbf{Y}_i \mid \mathbf{b}_i \sim N(\mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{Z}_i \cdot \mathbf{b}_i, \Sigma_i), \tag{10.4}$$
   $$\mathbf{b}_i \sim N(\mathbf{0}, D), \tag{10.5}$$

   where the same terminology is used as before (Section 3.3.1). Here, formula (10.4) refers to the defined hierarchical model of Section 10.2, while expression (10.5) denotes the prior density of the random effects.

   Based on formulation (10.4) - (10.5), the following marginal model, i.e., formula (3.24), and posterior distribution of the random effects follow (Verbeke and Molenberghs, 2000; Searle et al, 1996):

   $$\mathbf{Y}_i \sim N(\mathbf{X}_i \cdot \boldsymbol{\xi}, \mathbf{V}_i = \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' + \Sigma_i), \tag{10.6}$$
   $$\mathbf{b}_i \mid \mathbf{Y}_i \sim N\left[D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{Y}_i - \mathbf{X}_i \cdot \boldsymbol{\xi}), (\mathbf{Z}_i' \cdot \Sigma_i^{-1} \cdot \mathbf{Z}_i + D^{-1})^{-1}\right] \tag{10.7}$$

   Due to the definition of EB prediction for the normal random effects, i.e., formula (4.17), and expression (10.7), the EB estimation easily follows:

   $$\widehat{\mathbf{b}}_i = \mathrm{E}(\mathbf{b}_i \mid \mathbf{Y}_i) = D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{Y}_i - \mathbf{X}_i \cdot \boldsymbol{\xi}). \tag{10.8}$$

For the prediction of $\mathbf{Y}_i$, the outcome in (10.8) is plugged into the mean of the hierarchical model (10.4):

$$\widehat{\mathbf{Y}}_i = (\mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}) \cdot \mathbf{V}_i^{-1} \cdot \mathbf{y}_i + (\Sigma_i) \cdot \mathbf{V}_i^{-1} \cdot \mathbf{X}_i \cdot \boldsymbol{\xi}, \tag{10.9}$$

the familiar 'weighted average' of the observed outcomes $\mathbf{y}_i$ and the marginal mean $\mathbf{X}_i \cdot \boldsymbol{\xi}$.

2. *A first normal exponential version of LMM*

To show the arbitrariness of the posterior density, mentioned in Theorem 10.2.1, the normally distributed random effects is replaced by a vector of $n_i$ independent gamma random effects, where each outcome component $Y_{ij}$ is paired with a gamma random effect $g_{ij}$. The conventional density for a gamma variable $\phi$ is

$$f(\phi) = [\beta^{\alpha} \cdot \Gamma(\alpha)]^{-1} \cdot \phi^{\alpha-1} \cdot e^{-\phi/\beta}, \tag{10.10}$$

with parameters $\alpha, \beta \geq 0$. Equivalent to Verbeke and Molenberghs (2010), let $\alpha = 1$ and $\delta = 1/\beta$ in (10.10). Expression (10.10) is then expressed by

$$f(\phi) = \delta \cdot e^{-\phi \cdot \delta}, \tag{10.11}$$

i.e., the exponential density. Verbeke and Molenberghs (2010) motivated this choice to conveniently illustrate Theorem 10.2.1, in such a way that reasonably tractable closed-form expressions are provided, at the same time allowing for choice within the exponential framework. Here, a conditional density of the form (10.11) for $\phi = g_{ij}$ is chosen, with $\delta = \gamma_j \cdot y_{ij}$ and $\gamma_j$ is an unspecified parameter.

The marginal model (10.6) is retained, and coupled with the posterior density

$$f(\mathbf{g}_i \mid \mathbf{y}_i) = \prod_{j=1}^{n_i} \gamma_j \cdot y_{ij} \cdot e^{-g_{ij} \cdot \gamma_j \cdot y_{ij}}. \tag{10.12}$$

From expression (10.2), it immediately follows that the joint density of $\mathbf{y}_i$ and $\mathbf{g}_i$ equals to the product of the marginal density (10.6) and posterior density (10.12). By deriving some algebra (not presented in this thesis), the hierarchical and prior density follows as

$$f(\mathbf{y}_i \mid \mathbf{g}_i) = \frac{\left(\prod_{j=1}^{n_i} y_{ij}\right) \cdot e^{\boldsymbol{\theta}_i^{'} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\xi})} \cdot e^{\frac{1}{2} \cdot [(\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\xi})^{'} \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\xi}) + \boldsymbol{\theta}_i^{'} \cdot \mathbf{V}_i \cdot \boldsymbol{\theta}_i]}}{(2 \cdot \pi)^{n_i/2} \cdot \mid D \mid^{1/2} \cdot M_{n_i}(\mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{V}_i \cdot \boldsymbol{\theta}_i, \mathbf{V}_i)}, \tag{10.13}$$

$$f(\mathbf{g}_i) = \left(\prod_{j=1}^{n_i} \gamma_j\right) \cdot e^{(\mathbf{X}_i \cdot \boldsymbol{\xi})^{'} \cdot \boldsymbol{\theta}_i + \frac{1}{2} \cdot \boldsymbol{\theta}_i^{'} \cdot \mathbf{V}_i \cdot \boldsymbol{\theta}_i} \cdot M_{n_i}(\mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{V}_i \cdot \boldsymbol{\theta}_i, \mathbf{V}_i), \tag{10.14}$$

respectively. From the expressions above, $\boldsymbol{\theta}_i$ has components $\theta_{ij} = -g_{ij} \cdot \gamma_j$, and $M_n(\mathbf{k}, V) = E(Y_1 \ldots Y_n; \mathbf{k}, V)$, i.e., the sole $n$th order moment, relative to a normal distribution with mean $\mathbf{k}$ and variance $V$, each component occurs exactly ones.

From Willink (2005), Verbeke and Molenberghs (2010) pointed out that a simple recursive relationship is applicable to compute such moments, based on the concept of Hermite polynomials (Hildebrand, 1962):

$$M_n(\mathbf{k}, V) = k_n \cdot M_{n-1}(\mathbf{k}, V) + \sum_{j=1}^{n-1} v_{jn} \cdot M_{1,\ldots,j-1,j+1,\ldots,n-1}(\mathbf{k}, V), \tag{10.15}$$

The last term in (10.15) represents an $(n-2)$th order moment, with both the $j$th and $n$th components left out, while $k_j$ equals the $j$th element of vector $\mathbf{k}$ and $v_{jn}$ denotes the $(j,n)$th entry of the matrix $V$.

Furthermore, the EB and outcome estimations are expressed by

$$\widehat{g}_{ij} = \frac{1}{\gamma_j \cdot y_{ij}}, \tag{10.16}$$

$$\widehat{\mathbf{y}}_i = \frac{\mathbf{P}_{n_i}(\boldsymbol{\mu_i} - V_i \cdot \mathbf{z}_i, V_i)}{M_{n_i}(\boldsymbol{\mu_i} - V_i \cdot \mathbf{z}_i, V_i)}, \tag{10.17}$$

respectively, where $\mathbf{P}_{n_i}(\boldsymbol{\mu_i} - V_i \cdot \mathbf{z}_i, V_i)$ is an $n_i$-dimensional vector with components

$$P_{nj}(\mathbf{k}, V_i) = E(Y_1 \ldots Y_{i,j-1} Y_{i,j}^2 Y_{i,j+1} \ldots Y_n; \mathbf{k}, V), \tag{10.18}$$

and $\mathbf{z}_i$ equals to a vector with components $z_{ij} = 1/y_{ij}$. To derive the components of (10.18), Verbeke and Molenberghs (2010) used the following recursive relationship:

$$\begin{aligned} P_{nj}(\mathbf{k}, V_i) = k_j \cdot M_{n-1}(\mathbf{k}, V) + \sum_{k \neq j} v_{jk} \cdot E(Y_1 \ldots Y_{i,j-1} Y_{i,j}^2 Y_{i,j+1} \ldots Y_{i,k-1} Y_{i,k+1} \ldots Y_n) \\ + v_{jk} \cdot E(Y_1 \ldots Y_{i,j-1} Y_{i,j+1} \ldots Y_n). \end{aligned} \tag{10.19}$$

From these developments, Verbeke and Molenberghs (2010) denoted that an obvious consequence is present regarding the meaning of model parameters. More specifically, one might argue that there still is the hierarchical interpretation of (10.4)-(10.5) present in (10.6)-(10.7), since the combination of (10.6)-(10.7) is equivalent to the original hierarchical model (10.4)-(10.5), where parameters $\boldsymbol{\xi}$, $\Sigma_i$ and $D$ in general, are part of a hierarchical specification. The only difference is now that all three parameters $\boldsymbol{\xi}$, $\Sigma_i$ and $D$ occur in each of the two models (10.6)-(10.7), while these were separated in (10.4) and (10.5) by $\boldsymbol{\xi}$ and $\Sigma_i$, and $D$, respectively. However, authors like Verbeke and Molenberghs argued that a fundamental difference is present in parameter interpretation, even to the point of bearing on the inferences made, when one solely considers the marginal model (10.6) (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2007). To clarify this, Verbeke and Molenberghs (2010) referred to the model composed of (10.6) and (10.12), where all three parameters $\boldsymbol{\xi}$, $\Sigma_i$ and $D$ feature in the marginal model only and the hierarchical parameters, here $\gamma_j$, are full separated from the marginal ones. This implies that the hierarchical parameter is estimable only because it also is presented in marginal model (10.6), where is information in the data. In other words, it can be said that,

in the conventional hierarchical model, all parameters are identifiable from marginal model (10.6), i.e., the only one by which the data convey information. With these aspects in place, Verbeke and Molenberghs (2010) concluded that the model merely appears interpretable at a hierarchical, or enriched, level since (10.6) contains these, and only these parameters.

3. *A second normal exponential version of LMM*

Now, an alternative choice for $\delta$ is obtained in equation (10.11). Equivalent to Verbeke and Molenberghs (2010), the following choice is taken:

$$\delta = e^{\gamma_j \cdot y_{ij}}. \tag{10.20}$$

Using expansion

$$\prod_{j=1}^{n_i} e^{-q_{ij} \cdot e^{\gamma_j \cdot y_{ij}}} = \sum_{m_1=0}^{+\infty} \cdots \sum_{m_k=0}^{+\infty} \frac{(-q_{i1})^{m_1} \ldots (-q_{in_i})^{m_{n_i}}}{m_1! \ldots m_{n_i}!} \cdot e^{m_1 \cdot \gamma_1 \cdot y_{i1} + \cdots + m_{n_i} \cdot \gamma_{n_i} \cdot y_{in_i}}, \tag{10.21}$$

some basic algebra (not shown here) and obtaining the same terminology as in the first normal exponential case, the following hierarchical, prior, posterior model, EB and outcome estimations are derived:

$$f(\mathbf{y}_i \mid \mathbf{q}_i) = \frac{\prod_{j=1}^{n_i} e^{\gamma_j \cdot y_{ij}} \cdot e^{-q_{ij} \cdot e^{\gamma_j \cdot y_{ij}}} \cdot e^{-\boldsymbol{\mu}_i' \cdot \boldsymbol{\lambda}_m - \frac{1}{2} \cdot [(\mathbf{y}_i - \boldsymbol{\mu}_i)' \cdot V_i^{-1} \cdot (\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\lambda}_m' \cdot V_i \cdot \boldsymbol{\lambda}_m)]}}{(2 \cdot \pi)^{n_i/2} \cdot \mid V_i \mid^{1/2} \cdot \sum_{\mathbf{m}} \left( \prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right)}, \tag{10.22}$$

$$f(\mathbf{q}_i) = \sum_{\mathbf{m}} \left( \prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right) \cdot e^{\boldsymbol{\mu}_i' \cdot \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}_m' \cdot V_i \cdot \boldsymbol{\lambda}_m}, \tag{10.23}$$

$$f(\mathbf{q}_i \mid \mathbf{y}_i) = \prod_{j=1}^{n_i} e^{\gamma_j \cdot y_{ij}} \cdot e^{-q_{ij} \cdot e^{\gamma_j \cdot y_{ij}}} \tag{10.24}$$

$$\widehat{q}_{ij} = e^{-\gamma_j \cdot y_{ij}}, \tag{10.25}$$

$$\widehat{\mathbf{y}}_i = \frac{\sum_{\mathbf{m}} \left[ \prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j \cdot y_{ij}})^{m_j}}{m_j!} \right] \cdot e^{\boldsymbol{\mu}_i' \cdot \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}_m' \cdot V_i \cdot \boldsymbol{\lambda}_m} \cdot (\boldsymbol{\mu}_i + V_i \cdot \boldsymbol{\lambda}_m)}{\sum_{\mathbf{m}} \left[ \prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j \cdot y_{ij}})^{m_j}}{m_j!} \right] \cdot e^{\boldsymbol{\mu}_i' \cdot \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}_m' \cdot V_i \cdot \boldsymbol{\lambda}_m}}, \tag{10.26}$$

where $\mathbf{m}$ ranges over all non-negative integer vectors $\mathbf{m} = (m_1, \ldots, m_{n_i})$, and $\boldsymbol{\lambda}_m$ consist of components $\lambda_{mj} = (m_j + 1) \cdot \gamma_j$.

### 10.3.2 Exchangeable Data with Compound-Symmetry Covariance

In this part, a special but enlightening case of exchangeable, compound symmetry data is considered, meaning that all members of a cluster have the same mean $\mu_i$ and a compound-symmetry structure for the variance-covariance matrix, i.e., $V_i = \sigma^2 \cdot I_{n_i} + d \cdot J_{n_i}$, where $I_{n_i}$ is an $n_i$-dimensional identity matrix and $J_{n_i}$ is an $n_i$ x $n_i$ matrix consisting of 1's.

Similar to Section 10.3.1, this section will be sub-divided into three parts, i.e., expressing the hierarchical, prior, posterior model, EB and outcome estimations for (1) the standard LMM, (2)a first normal exponential version of LMM and (3) a second normal exponential version of LMM. Since this chapter only focuses on the principle idea behind enriched-data problems and essential non-identifiability, no attention will be given on the derivations, but mainly on the expressions.

1. *The standard LMM*

   To express the hierarchical, prior, posterior model, EB and outcome estimations of the standard LMM, Verbeke and Molenberghs (2010) first derived the following two main expressions:

$$V_i^{-1} = \frac{1}{\sigma^2} \cdot \left( I_{n_i} - \frac{d}{d \cdot n_i + \sigma^2} \cdot J_{n_i} \right), \tag{10.27}$$

$$\mid V_i \mid = \sigma^{2 \cdot n_i} + n_i \cdot \sigma^{2 \cdot (n_i - 1)} \cdot d. \tag{10.28}$$

   Imputing both expressions (10.27) and (10.28) in their derivations, Verbeke and Molenberghs (2010) formulated the hierarchical, prior, marginal and posterior model as

$$\mathbf{Y}_i \mid b_i \sim N(\mathbf{1}_{n_i} \cdot \mu_i + \mathbf{1}_{n_i} \cdot b_i, \sigma^2 \cdot I_{n_i}), \tag{10.29}$$

$$b_i \sim N\left(0, d\right), \tag{10.30}$$

$$\mathbf{Y}_i \sim N(\mathbf{1}_{n_i} \cdot \mu_i, V_i = \sigma^2 \cdot I_{n_i} + d \cdot J_{n_i}), \tag{10.31}$$

$$b_i \mid \mathbf{Y}_i \sim N\left[ \frac{d \cdot n_i}{d \cdot n_i + \sigma^2} \cdot (\bar{y}_i - \mu_i), \frac{\sigma^2}{d \cdot n_i + \sigma^2} \cdot d \right], \tag{10.32}$$

   respectively, where $\mathbf{1}_{n_i}$ denotes a $n_i$-vector of ones and $\bar{y}_i$ be the average of the components of outcome vector $\mathbf{y}_i$.

   The EB and outcome estimations are given by

$$\widehat{b}_i = \frac{d \cdot n_i}{d \cdot n_i + \sigma^2} \cdot (\bar{y}_i - \mu_i), \tag{10.33}$$

$$\widehat{\mathbf{Y}}_i = \frac{d \cdot n_i \cdot \bar{y}_i + \sigma^2 \cdot \mu_i}{d \cdot n_i + \sigma^2} \cdot \mathbf{1}_{n_i}, \tag{10.34}$$

   respectively.

2. *A first normal exponential version of LMM*

   Like in the first normal exponential version of LMM in Section 10.3.1, where posterior density (10.12) is coupled with marginal model (10.6), the marginal model (10.31) will be coupled with posterior density

$$f(g_i \mid \mathbf{y}_i) = \gamma \cdot \bar{y}_i \cdot e^{-g_i \cdot \gamma \cdot \bar{y}_i}, \tag{10.35}$$

leading to the following prior and hierarchical model respectively:

$$f(g_i) = \gamma \cdot e^{-g_i \cdot \mu_i \cdot \gamma + \frac{1}{2} \cdot \frac{g_i^2 \cdot \gamma^2}{n_i} \cdot (d \cdot n_i + \sigma^2)} \cdot \left[ \frac{n_i \cdot \mu_i - g_i \cdot \gamma \cdot (d \cdot n_i + \sigma^2)}{n_i} \right], \tag{10.36}$$

$$f(\mathbf{y}_i \mid g_i) = \frac{n_i \cdot \bar{y}_i \cdot e^{-\frac{1}{2} \cdot \left[ \frac{1}{\sigma^2} \cdot (\mathbf{y}_i - \mathbf{1}_{n_i} \cdot \bar{y}_i)' \cdot (\mathbf{y}_i - \mathbf{1}_{n_i} \cdot \bar{y}_i) + \frac{n_i}{d \cdot n_i + \sigma^2} \cdot (\bar{y}_i - \mu_i)^2 \right] - g_i \cdot \gamma \cdot (\bar{y}_i - \mu_i)}}{(2 \cdot \pi)^{n_i/2} \mid V_i \mid^{1/2} \cdot e^{\frac{1}{2} \cdot \frac{g_i^2 \cdot \gamma^2}{n_i} \cdot (d \cdot n_i + \sigma^2)} \cdot [n_i \cdot \mu_i - g_i \cdot \gamma \cdot (d \cdot n_i + \sigma^2)]}. \tag{10.37}$$

Furthermore, the EB and outcome predictions are expressed by

$$\widehat{g}_i = \frac{1}{(\gamma \cdot \bar{y}_i)}, \tag{10.38}$$

$$\widehat{\mathbf{y}}_i = \frac{\left\{ \left[ n_i \cdot \mu_i - \frac{1}{\bar{y}_i} \cdot (d \cdot n_i + \sigma^2) \right]^2 + n_i \cdot (d \cdot n_i + \sigma^2) \right\} \cdot \mathbf{1}_{n_i}}{n_i \cdot \left[ n_i \cdot \mu_i - \frac{1}{\bar{y}_i} \cdot (d \cdot n_i + \sigma^2) \right]}, \tag{10.39}$$

respectively.

3. *A second normal exponential version of LMM*

At last, marginal model (10.31) will be coupled with posterior density

$$f(q_i \mid \mathbf{y}_i) = e^{\gamma \cdot \bar{y}_i - q_i \cdot e^{\gamma \cdot \bar{y}_i}}. \tag{10.40}$$

Th following expressions are then obtained for hierarchical and prior model, respectively:

$$f(\mathbf{y}_i \mid q_i) = \frac{e^{-\frac{1}{2} \cdot \left[ \frac{1}{\sigma^2} \cdot (\mathbf{y}_i - \mathbf{1}_{n_i} \cdot \bar{y}_i)' \cdot (\mathbf{y}_i - \mathbf{1}_{n_i} \cdot \bar{y}_i) + \frac{n_i}{d \cdot n_i + \sigma^2} \cdot (\bar{y}_i - \mu_i)^2 \right] + \gamma \cdot \bar{y}_i - q_i \cdot e^{\gamma \cdot \bar{y}_i}}}{(2 \cdot \pi)^{n_i/2} \mid V_i \mid^{1/2} \cdot \sum_{m=0}^{\infty} \frac{(-q_i)^m}{m!} \cdot e^{\mu_i \cdot \gamma \cdot (m+1) + \frac{1}{2} \cdot \frac{\gamma^2 \cdot (m+1)^2}{n_i} \cdot (d \cdot n_i + \sigma^2)}}, \tag{10.41}$$

$$f(q_i) = \sum_{m=0}^{\infty} \frac{(-q_i)^m}{m!} \cdot e^{\mu_i \cdot \gamma \cdot (m+1) + \frac{1}{2} \cdot \frac{\gamma^2 \cdot (m+1)^2}{n_i} \cdot (d \cdot n_i + \sigma^2)}. \tag{10.42}$$

For the EB and outcome predictions, the following formulas hold:

$$\widehat{q}_i = e^{-\gamma \cdot \bar{y}_i}, \tag{10.43}$$

$$\widehat{\mathbf{y}}_i = \frac{\sum_{m=0}^{\infty} \frac{(e^{-\gamma \cdot \bar{y}_i})^m}{m!} \cdot e^{\mu_i \cdot \gamma \cdot (m+1) + \frac{1}{2} \cdot \frac{\gamma^2 \cdot (m+1)^2}{n_i} \cdot (d \cdot n_i + \sigma^2)} \cdot \left[ \mu_i + \frac{\gamma \cdot (m+1)}{n_i} \cdot (d \cdot n_i + \sigma^2) \right] \cdot \mathbf{1}_{n_i}}{\sum_{m=0}^{\infty} \frac{(e^{-\gamma \cdot \bar{y}_i})^m}{m!} \cdot e^{\mu_i \cdot \gamma \cdot (m+1) + \frac{1}{2} \cdot \frac{\gamma^2 \cdot (m+1)^2}{n_i} \cdot (d \cdot n_i + \sigma^2)}}, \tag{10.44}$$

# Chapter 11

# Analyzing the Recurrent Asthma Attacks in Children

Until now, only a theoretical discussion has been provided for the WGN model (4.5)-(4.8). The asthma data of Chapter 2 will be analyzed with the proposed WGN model (4.5)-(4.8). Specifically, the exponential model will be used here to investigate the modeling framework on the data. Descriptive statistics of the dataset can be found in Appendix G.1.

Let $Time_{ij}$ represents the time to recurrence of an asthma attack, and assume the same terminology as in Section 4.1. Then, $Time_{ij}$ will be modeled with the following (random-intercept) WGN model:

$$Time_{ij} \mid \mathbf{b}_i, \theta_{ij} \sim \text{Weibull}(1, k_{ij}), \tag{11.1}$$

$$k_{ij} = \theta_{ij} \cdot e^{\xi_0 + \xi_1 \cdot T_i + b_i}, \tag{11.2}$$

$$b_i \sim N(0, d), \tag{11.3}$$

$$\theta_{ij} \sim \text{Gamma}(\alpha, 1/\alpha), \tag{11.4}$$

where $T_i = 0$ if patient $i$ got a placebo and 1 if patient $i$ got the drug.

The remainder of this chapter is organized as follows.
In Section 11.1, different hierarchical modeling choices will be explored for (11.1)-(11.4), i.e., (1) the traditional Exponential model (without gamma and normal random effect), (2) the Exponential-gamma model (without normal random effect), (3) the Exponential-normal model (without Gamma random effect) and (4) the Exponential-gamma-normal model. Full likelihood estimation (Section 4.2.1) is used to explore all models. Next, focusing on the combined model (referred as (4)) and due to the contribution of Molenberghs et al (2014), the pairwise likelihood (Section 4.2.2) principle is compared with the full likelihood, both when taking censoring into account or not. Section 11.2, on the other hand, focuses on the comparison between the hierarchical and its marginalized combined model, both fitted with full and pairwise likelihood. Since Section 11.1 restricted the modeling framework to a random-intercept approach, Section 11.2 will provide a discussion on the (more extended) random-slope approach, where formula (11.2) is replaced by $k_{ij} = \theta_{ij} \cdot e^{\xi_0 + (\xi_1 + b_{2i}) \cdot T_i + b_{1i}}$. Both random effects are assumed to be independently normally distributed. Additionally, Molenberghs et al (2014) and Efendi et al (2014) extended this research by performing a simulation study to evaluate the performance of the hierarchical

model, fitted under full likelihood and pairwise likelihood, and of the hierarchical and its marginalized combined model, respectively. A brief discussion of their additional studies can be found in Molenberghs et al (2014) and Efendi et al (2014). While Section 11.3 discusses the results performed from the gradient function to assess fit of the random effects distribution, local influence diagnostics are discussed in Section 11.4. To conclude, a brief discussion of non-identifiability problems is provided.

In this chapter, statistical software package SAS is used to perform the analysis. All code can be found in Appendix G.2.

## 11.1 Several Hierarchical Modeling Strategies with different Estimation Strategies

By using the "NLMIXED" procedure in SAS, all four models, i.e., (1) the traditional Exponential model (without gamma and normal random effect), (2) the Exponential-gamma model (without normal random effect), (3) the Exponential-normal model (without Gamma random effect) and (4) the Exponential-gamma-normal model, can easily be fitted for the asthma data. Results of the fitted models are listed in the Table 11.1.

**Table 11.1:** *Parameter estimates and standard errors for the regression coefficients in* (1) *the Exponential model,* (2) *Exponential-gamma model,* (3) *Exponential-normal model and* (4) *Exponential-gamma-normal model, also referred as the CM. Maximum likelihood estimation with partial marginalization (Section 4.2.1) was done.*

| | | **Exponential** | **Exponential-gamma** |
|---|---|---|---|
| Effect | Parameter | Estimate (s.e.) | Estimate (s.e.) |
| Intercept | $\xi_0$ | $-3.3709$ (0.0772) | $-3.9782$ (15.354) |
| Treatment effect | $\xi_1$ | $-0.0726$ (0.0475) | $-0.0755$ (0.0605) |
| Shape parameter | $\lambda$ | 0.8140 (0.0149) | 1.0490 (16.106) |
| Std. dev. random effect | $\sqrt{d}$ | $--$ | $--$ |
| Gamma parameter | $\alpha$ | $--$ | 3.3192 (0.3885) |
| -2 log-likelihood | | 18693 | 18715 |
| | | **Exponential-normal** | **Combined** |
| Effect | Parameter | Estimate (s.e.) | Estimate (s.e.) |
| Intercept | $\xi_0$ | $-3.8095$ (0.1028) | $-3.9923$ (20.337) |
| Treatment effect | $\xi_1$ | $-0.0825$ (0.0731) | $-0.0887$ (0.0842) |
| Shape parameter | $\lambda$ | 0.8882 (0.0180) | 0.8130 (16.535) |
| Std. dev. random effect | $\sqrt{d}$ | 0.4097 (0.0386) | 0.4720 (0.0416) |
| Gamma parameter | $\alpha$ | $--$ | 6.8414 (1.7146) |
| -2 log-likelihood | | 18611 | 18629 |

Since main interest of researchers lies in the exploration of the overall treatment effect on patients, a formal assessment of the treatment effect $\xi_1$ is explored for all four models (Table 11.2). Similar results are obtained for all four models, in such a way that the treatment effects are similar in strength. Including both random effects, however, reduces the evidence, relative to the exponential model. However, too parsimonious an association structure might lead to liberal test behaviour.

**Table 11.2:** *Wald test for the assessment of treatment effect.*

| Model | $Z$-value | $p$-value |
|---|---|---|
| Exponential | $-1.5283$ | 0.1264 |
| Exponential-gamma | $-1.2480$ | 0.2120 |
| Exponential-normal | $-1.1293$ | 0.2588 |
| Exponential-gamma-normal | $-1.0534$ | 0.2921 |

Similar to the maximum likelihood principle with partial marginalization (Section 4.2.1) above, pairwise likelihood (Section 4.2.2) is considered to model the Exponential-gamma-normal model, with proper inclusion of the censored observations. Since the SAS procedure "NLMIXED" does not contain pairwise likelihood standardly, a SAS macro was developed by Molenberghs et al (2014), in conjunction with the "NLMIXED" procedure, to perform pairwise likelihood (Appendix G.2). Results of this particular estimation technique, both incorporating with and without censoring, can be found in Table 11.3.

**Table 11.3:** *The Exponential-gamma-normal model fitted with and without censoring. Both maximum likelihood with partial marginalization and pairwise likelihood estimation was done. (model-based s.e.; empirically corrected s.e.)*

| | | **Full likelihood** | **Pairwise likelihood** |
|---|---|---|---|
| Effect | Parameter | Estimate (s.e.) | Estimate (s.e.) |
| Without censoring | | | |
| Intercept | $\xi_0$ | $-3.9923$ (20.337) | $-3.4862$ (6.2316; 0.0856) |
| Treatment effect | $\xi_1$ | $-0.0887$ (0.0842) | $-0.1060$ (0.0203; 0.0953) |
| Shape parameter | $\lambda$ | 0.8130 (16.534) | 0.8272 (5.1551; 0.0049) |
| Std. dev. random effect | $\sqrt{d}$ | 0.4720 (0.0416) | 0.3958 (0.0202; 0.0383) |
| Gamma parameter | $\alpha$ | 6.8414 (1.7146) | 6.7758 (0.6648; 1.1875) |
| With censoring | | | |
| Intercept | $\xi_0$ | $-4.0195$ (28.663) | $-3.6233$ (0.4998; 0.09381) |
| Treatment effect | $\xi_1$ | $-0.1115$ (0.0996) | $-0.1269$ (0.0221; 0.10571) |
| Shape parameter | $\lambda$ | 0.7882 (22.592) | 0.9189 (0.4590; 0.0003) |
| Std. dev. random effect | $\sqrt{d}$ | 0.5620 (0.0506) | 0.4443 (0.0211; 0.03906) |
| Gamma parameter | $\alpha$ | 3.5633 (0.6281) | 4.5882 (0.3627; 0.71248) |

When full likelihood is performed, results in Table 11.3 indicate a presence of overdispersion, regardless of whether censoring is taken into account in the model. However, since the standard errors are far from plausible, difficulties with convergence can be present. This possible issue does not seem to occur in the pairwise likelihood estimation technique, but with the additional observation that overdispersion now disappears.

To address this issue even further, Molenberghs et al (2014) refitted all four models of Table 11.3, i.e., the Exponential-gamma-normal model (1) with incorporating censoring and fitted with full likelihood, (2) with incorporating censoring and fitted with pairwise likelihood, (3) without incorporating censoring and fitted with full likelihood, and (4) without incorporating censoring and fitted with pairwise likelihood, with the additional

constrain of setting $\lambda$ equally to one. This approach is adopted here, where the obtained results are presented in Table 11.4.

**Table 11.4:** *Parameter estimates and standard errors for the regression coefficients in the hierarchical combined model with $\lambda = 1$. Incorporation was done with and without censoring . Estimation was proceeded with both maximum likelihood with partial marginalization and pairwise likelihood. (model-based s.e.; empirically corrected s.e.)*

| | | **Full likelihood** | **Pairwise likelihood** |
|---|---|---|---|
| Effect | Parameter | Estimate (s.e.) | Estimate (s.e.) |
| Without censoring | | | |
| Intercept | $\xi_0$ | $-4.1993$ (0.0713) | $-3.6758$ (0.0176; 0.0869) |
| Treatment effect | $\xi_1$ | $-0.0887$ (0.0842) | $-0.1060$ (0.0203; 0.0953) |
| Std. dev. random effect | $\sqrt{d}$ | 0.4721 (0.0416) | 0.3958 (0.0202; 0.0383) |
| Gamma parameter | $\alpha$ | 6.8410 (1.7144) | 6.7754 (0.6648; 1.1874) |
| With censoring | | | |
| Intercept | $\xi_0$ | $-4.2575$ (0.0833) | $-3.7072$ (0.0160; 0.0875) |
| Treatment effect | $\xi_1$ | $-0.1116$ (0.0996) | $-0.1267$ (0.0218; 0.1122) |
| Std. dev. random effect | $\sqrt{d}$ | 0.5620 (0.0506) | 0.4446 (0.0177; 0.0424) |
| Gamma parameter | $\alpha$ | 3.5634 (0.6282) | 4.5833 (0.1747; 0.1895) |

Contrary to the fitted results of Table 5, no disparity is present in the overdispersion results and the standard errors are plausible throughout. Equivalent to the previous discussed analyses (Table 11.2), the assessment of treatment effect in the combined model is explored for all four models (Table 11.5), indicating similar results as before (Table 11.2).

**Table 11.5:** *Wald test for the assessment of treatments effect in the Exponential-gamma-normal model.*

| Model | $Z$-value | $p$-value |
|---|---|---|
| Full Likelihood | | |
| Without censoring | $-1.0534$ | (0.1461) |
| With censoring | $-1.1205$ | (0.1312) |
| Pairwise Likelihood | | |
| Without censoring | $-1.1123$ | (0.1330) |
| With censoring | $-1.1292$ | (0.1294) |

A few comments are in place.
First, similar to the findings of Molenberghs et al (2014), a faster convergence time was reached with pseudo-likelihood, when comparing it to the full likelihood methodology. Second, Molenberghs et al (2014) pointed out that pseudo-likelihood is more robust against the choice of starting values. Intuitively, this seems reasonable, since the computational behavior pattern of pairwise likelihood principle is comparable when analyzing bivariate data. The higher the order of the likelihood, the more vulnerable to numerical instabilities.

To conclude the discussion on hierarchical modeling for the asthma dataset, a choice of model is made based on the following three aspects: (1) convergence presence, (2) best accounting for censoring and (3) redundant presence of the shape parameter. As result, the exponential-gamma-normal model with incorporating censoring and fitted with full likelihood is chosen as final one. However, once should be careful not to over-interpret the results derived from a couple of data analyses. Therefore, Molenberghs et al (2014) additionally performed a simulation study to evaluate the exponential-gamma-normal under full likelihood and pairwise likelihood. Specifically, different settings were considered, in order to explore the impact of sample size, censoring percentage, and estimation method. The following sample sizes, censoring percentages and estimation methods were considered:

1. Sample sizes consisting of 50, 100 and 200 subjects;
2. A total of 10%, 25% and 50% of the observations within a subject are censored;
3. Full likelihood and pseudo-likelihood.

For each different setting (9 in total), 500 datasets were generated. Since this thesis does not cover simulation studies, its relevant to present the concluding remarks of Molenberghs et al's (2014) simulation study. First, the proportion of non-converging increased with censoring. Secondly, pseudo-likelihood reduces consistency relative to the full likelihood principle. Thirdly, with increasing censoring percentage and under full likelihood, loss of some consistency will be present of the estimates of the exponential-gamma-normal model. A detailed discussion of the simulation study can be found in Molenberghs et al (2014).

## 11.2 The Combined Model and its Marginalized Version

Like mentioned before, Section 11.1 mainly focuses on the random-intercept approach. In this section, attention will be given on the random-slope approach, where the same estimation strategies, i.e., full likelihood and pairwise likelihood, are considered to explore the hierarchical exponential-gamma-normal model (with incorporation of censoring) and its marginalized version (Section 5.1). In the normal random effects structure, let $b_{1i}$ the random intercept with variance $d_1$ and denote $b_{2i}$ the random slope with variance $d_2$ for subject $i$, $i = 1, \ldots, 232$. Both random effects are assumed to be independent. Results of the fitted models are listed in Table 11.6.

From Table 11.6, a few remarks are drawn. Looking only at the full likelihood outcomes, similar results were achieved between the hierarchical and marginalized combined model. Contrary to Section 11.1, these models include two normally distributed random effects. Therefore, connector function (5.5) uses a different vector $\mathbf{z}_{ij}$, implying a change in the treatment effect estimate upon the marginalization. However, as can be observed, an extreme change is not present here. Furthermore, likelihood ratios are invariant (Griswold and Zener, 2004), since marginalization does not change the likelihood.

**Table 11.6:** *Parameter estimates and standard errors for the regression coefficients in the hierarchical and its marginalized combined model, with censoring. Estimation was done with both maximum likelihood with partial marginalization and pairwise likelihood. (model-based s.e.; empirically corrected s.e.)*

| | | Hierarchical Combined | | Marginalized Combined | |
|---|---|---|---|---|---|
| | | Full likelihood | Pairwise likelihood | Full likelihood | Pairwise likelihood |
| Effect | Parameter | Estimate (s.e.) | Estimate (s.e.) | Estimate (s.e.) | Estimate (s.e.) |
| Treatment effect | $\xi_1$ | $-0.113$ (0.106) | $-0.127$ (0.105) | $-0.111$ (0.102) | $-0.127$ (0.105) |
| Shape parameter | $\lambda$ | 0.014 (0.001) | 0.025 (0.002) | 0.017 (0.001) | 0.027 (0.003) |
| Std. dev. random effect | $\sqrt{d_1}$ | 0.560 (0.068) | 0.445 (0.039) | 0.560 (0.068) | 0.445 (0.039) |
| Std. dev. random effect | $\sqrt{d_2}$ | 0.077 (0.734) | $11\text{E}-4$ ($11\text{E}-4$) | 0.077 (0.741) | $20\text{E}-6$ ($20\text{E}-6$) |
| Gamma parameter | $\alpha$ | 3.566 (0.632) | 4.583 (0.708) | 3.566 (0.632) | 4.584 (0.708) |
| -2 log-likelihood | | 16649 | | 16649 | |

In the pairwise likelihood results, equivalent to the full likelihood ones, similar results were obtained between both models. Even though the estimated random slope variance is approximately zero for all models, this is more pronounced in the pairwise-likelihood results. However, this does not contradict the results from the full likelihood, where the estimated random slope variance was non significant. The treatment effect is not significant in all models!

While Section 11.1 did not provided any discussion on the conjugate random effect parameter $\alpha$, a brief discussion of the estimate will be given. Exploring the standard errors of all four models in Table 11.6, a significant conjugate random effect parameter is present. Thus, when fitting all models, overdispersion seems to be present in the data, indicating that the need to account for overdispersion in the model is necessary to adequately fit the data.

From a computational point of view, no convergence problems arise when fitting all models. To conclude the discussion on marginal modeling, it should be mentioned that once should be careful not to over-interpret the results (similar to the hierarchical modeling strategies of Section 11.1). A similar simulation study from Molenberghs et al (2014) was constructed by Efendi et al (2014), but now evaluating the performance of the hierarchical combined model and its marginalized version. The following sample sizes, censoring percentages ad estimation methods were considered:

1. Sample sizes consisting of 20, 40, 60 and 80 subjects, with 10 observations per subject;

2. A total of 0%, 10%, 25% and 50% of the observations within a subject are censored;

3. Full likelihood and pseudo-likelihood.

For each different setting (16 in total), 500 replicants were generated. From Efendi et al's (2014) study, a few conclusions were found. Similar to Molenberghs et al (2014), the percentage of non-convergence increased with censoring, but reduces with higher number of sample sizes. Pairwise likelihood had a little beneficial impact on convergence, at the cost of increased computation time. Computational time increased with sample size, as well as with censoring. From a estimation point of view, relative bias was rather insensitive

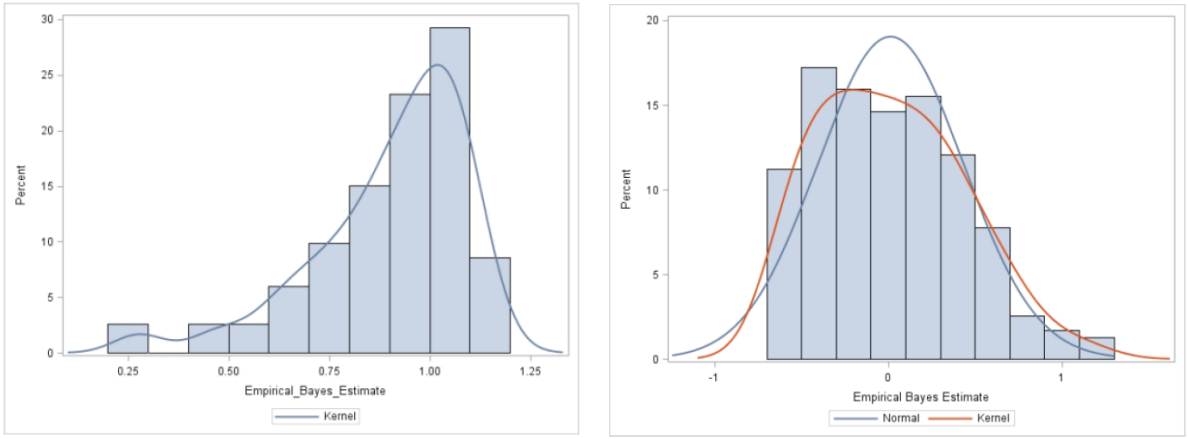


**Figure 11.1:** *Distribution of the EB estimation of the conjugate (left) and normal random effect (right). The underlying kernel densities for both distributions are plotted as well.*

to the estimation method. The model based and Monte Carlo standard errors are quite similar, as would expected. A full detailed discussion of the study can be found in Efendi et al (2014).

## 11.3 Gradient Function for Assessing Fit of Random-Effects Distribution

In Sections 11.1 and 11.2, no attention was given to the EB estimation of the conjugate random effect and normal random effect, explored in Section 4.3.2 and 4.3.1, respectively. Therefore, this section places emphasis on the EB estimation of the conjugate and normal random effect in the exponential-gamma-normal model, defined in Section 11.1 (random-intercept approach). Moreover, the gradient function of Section 8.2 is used as graphical exploratory diagnostic tool to assess whether the assumed random-effects distribution produces an adequate fit to the data, in terms of marginal likelihood. In particular, this is done for the normal random effect (11.3). In case of model misspecification, the gradient function gives an important, albeit informal, indication on how the model can be improved in terms of random-effects distribution.

To compute the EB estimation of both random effects, different strategies are obtained. For the normal random effect, no extra procedures have to be written in SAS, since the SAS procedure "NLMIXED" already implemented a standard output for EB estimation. For the gamma random effects, a modified version of Iddi et al's (2014) macro is used to derive the EB estimation. Results of both EB estimations are found in Figure 11.1.
To achieve the gradient function of the normal random effect, related to model (11.1)-(11.4), a modified version of the SAS code from Verbeke and Molenberghs (2013) is used. The resulted gradient function is shown in Figure 11.2. Also indicated by Figure 11.1 (right), the gradient function indicates that the model can be improved in terms of likelihood by moving probability mass from the region [-1.2; -0.6] toward the region [-0.4; -0.2].

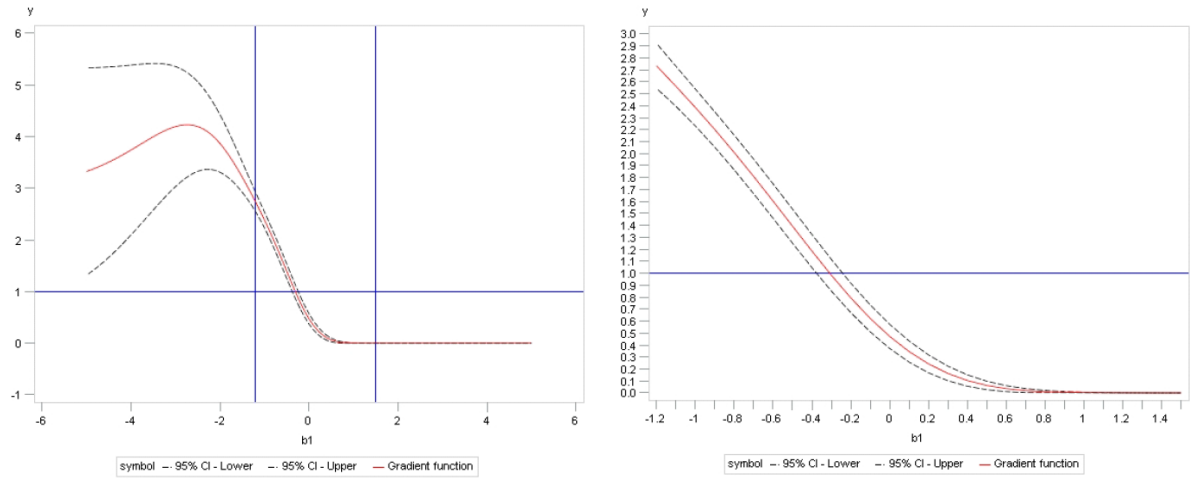


**Figure 11.2:** *Gradient function and 95% pointwise confidence bands for the hierarchical exponential-gamma-normal model, where, conveniently, normal distribution is assumed as a distribution for random intercepts.*

One way to improve the model is by replacing the normality assumption for the random intercept $b_i$ by a flexible distribution that captures most distributions quite well. A possible good choice, proposed by Verbeke and Lesaffre (1996) and Verbeke and Molenberghs (2000), is a finite mixture of normals, as they can handle skewness as well as symmetry, unimodality as well multimodality. By looking at the kernel density plot in Figure 11.1 (right), it seems reasonable to assume the following combined model:

$$Time_{ij} \mid \mathbf{b}_i, \theta_{ij} \sim \text{Weibull}(1, k_{ij}), \tag{11.5}$$

$$k_{ij} = \theta_{ij} \cdot e^{\xi_0 + \xi_1 \cdot T_i + b_i}, \tag{11.6}$$

$$b_i \sim \pi_1 \cdot N(\mu_1, d_1) + \pi_2 \cdot N(\mu_2, d_2), \tag{11.7}$$

$$\theta_{ij} \sim \text{Gamma}(\alpha, 1/\alpha), \tag{11.8}$$

with $\pi_1 + \pi_2 = 1$. Additionally, Verbeke and Molenberghs (2013) pointed out that the assumption $\text{E}(b_i)$ is imposed by the restriction $\pi_1 \cdot \mu_1 + \pi_2 \cdot \mu_2 = 0$. Intuitively, by looking at the kernel density plot in Figure 11.1 (right), an approximate estimation can be made for some parameters of the random intercept distribution (11.7), i.e., $\mu_1 \approx -0.2$, $\mu_2 \approx 0.1$ and $d_1 < d_2$.

To end the discussion on the gradient function, it should be noted that test-statistic (8.5) can be used to check these conclusions. Here, a significance outcome is expected for non-normality, and a significant one is expected for the finite mixture of normals (11.7).

## 11.4 Local Influence Analysis

To explore the behavior pattern on recurrent asthma attacks of patients, local influence diagnostics is advised. Therefore, the local influence paradigm of Rakhmawati, Molenberghs, Verbeke and Faes (2014) is used to detect outlying behavior patterns in the exponential-normal (EN) and exponential-gamma-normal (EGN) model of Section 11.1

(random-intercept approach). Index plots (versus patient ID's) for various local influence analysis are given in Figure 11.3. The top row of the plot represents the total local influence, with subsequent rows representing influence for the following sub-vectors: fixed effects $\{\xi_0, \xi_1\}$, shape parameter $\lambda$, random-intercept variance $d$, and, for the EGN model, overdispersion parameter $\alpha$, respectively. Patient # 182 stand out in the EGN with large total influence $C_{182}$, when compared to other patients. However, compared to the EN model, no large total influence $C_{182}$ was observed there. Overall, no outlying influential patients were observed in the EN, while only one (Patient # 182) was indicated by the EGN. Its therefore important to explore underlying reasons for observed local influence $C_{182}$ in the EGN. By looking at the sub-vectors, no large influence on $\{\xi_0, \xi_1\}$, $\lambda$ and $d$ were observed for # 182, whereas a large influence on overdispersion parameter $\alpha$ was obtained for # 182. In other words, by adding an overdispersion parameter to the EN model, resulting in the EGN model, opportunities are created that specific subjects (here patient # 182) will affect that parameter. Adding such a parameter also influence other patients in their estimates. For example, patient # 127 imputed a large influence on shape parameter $\lambda$, while # 7 and # 69, for example, indicate a large influence on $d$ in the EGN model. The interpretable components do not lead to additional insight (Figure 11.4).

## 11.5 Discussion on non-identifiability problems

To end the discussion of the performed asthma dataset analysis, a brief discussion is provided about non-identifiability problems in the asthma dataset. To do so, reference is made to Molenberghs, Njagi, Kenward, and Verbeke (2012), where, for purpose of illustration, a subset of data is considered not consisting of censored observations, i.e., data points for which the corresponding "end-to-observation" period corresponds to an attack. Since these authors uses a parametric proportional hazards Weibull-Gamma frailty model, abstraction is made from the proposed exponential-gamma-normal model (Chapter 11) in such a way that this section is only used informally.

In their study, a defined Gamma posterior is replaced with a normal one. As the marginal fit remains the same for both choices, different predictions were observed for the conditional hazard. In the Gamma choice, a prediction was created which lies much closer to the population hazard than the normal choice. Therefore, Molenberghs, Njagi, Kenward, and Verbeke (2012) showed that their disparate inferences occur in disturbing conjunction with an unaltered marginal model in the asthma dataset.

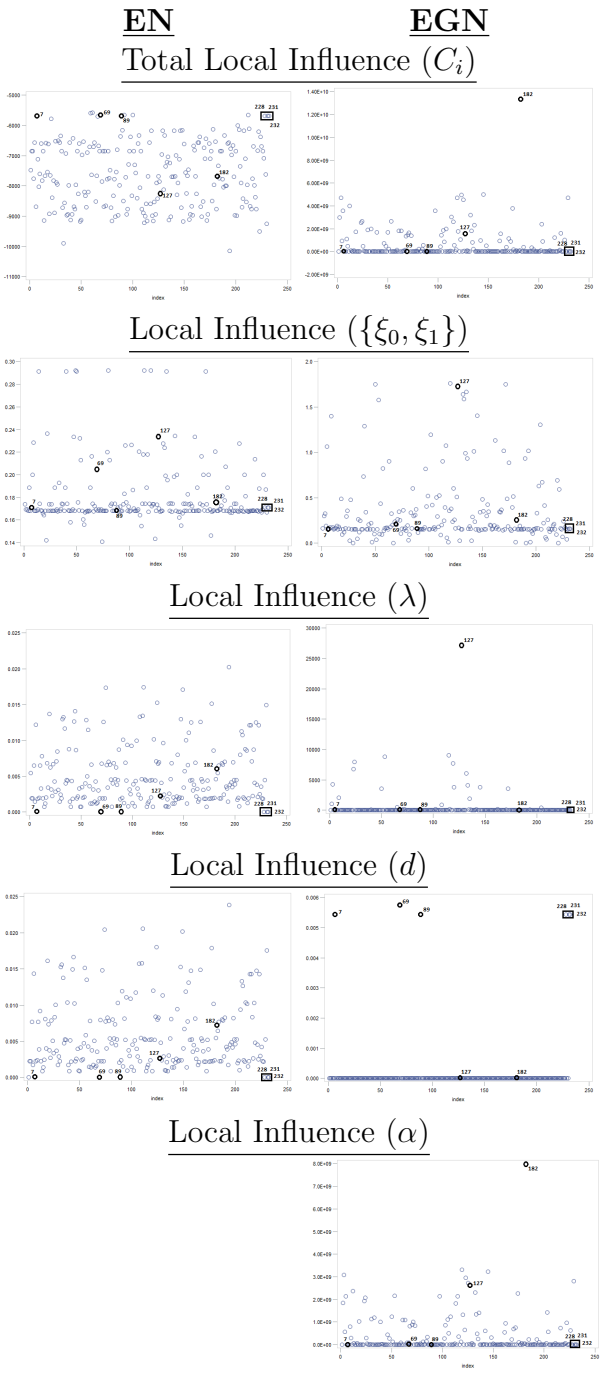


**Figure 11.3:** *Asthma Data. Local Influence plots*

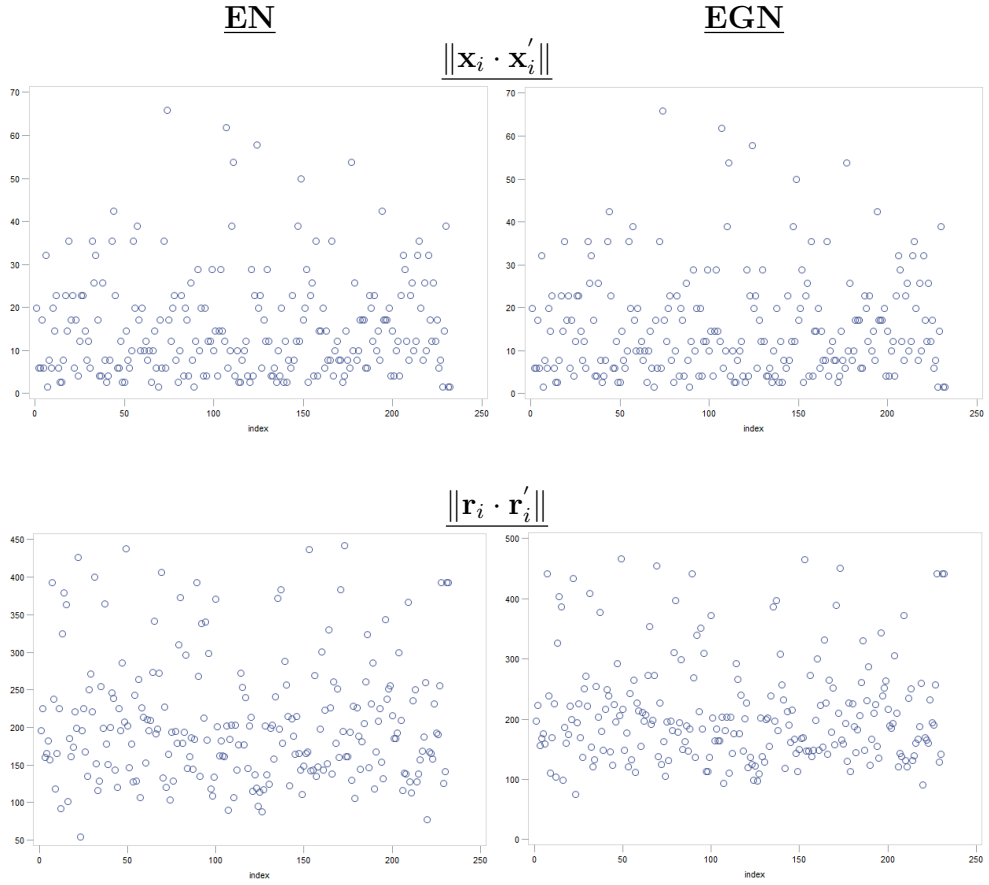


**Figure 11.4:** *Asthma Data. Plot of interpretable components of local influence*

# Chapter 12

# Conclusion

The general objective of the thesis was to propose flexible statistical methods to overdispersed, hierarchical data. Specifically, full attention was given to the WGN model of Molenberghs et al (2010), that fully accounts for overdispersion with conjugate random effects and takes the hierarchical structure into account by normal random effects. Due to the 'new' existence of the framework, a good amount of research is done nowadays, leading to new methodological insights. The structure of the thesis is organized as follows: (1) a theoretical discussion about the framework and (2) a practical application were the methodology is applied to. In the former one, special attention is provided on the developed mathematical derivations. For the latter one, the asthma dataset is used to show the convenient improvements of fitting, when comparing it with classical methodologies.

In Chapter 3 and 4, a broad and extensive discussion is outlined for the proposed CM framework of Molenberghs et al (2010). As starting point, a general overview is provided of the standard GLM framework. Apart from taking into account correlation induced by repeated measurements from the same cluster (subject) and the association between the different outcomes with normal random effects, the often-restrictive mean-variance prescription in the model for the non-Gaussian outcome has explicitly been addressed, and taken into account by the inclusion of conjugate random effects. Estimation strategies like maximum likelihood with partial marginalization and pairwise likelihood has been suggested as two main techniques to obtain parameter fits. As pointed out, both estimation strategies can easily be performed in standard statistical software packages. Due to the strong conjugacy principle, the WGN model is explored in detail to model time-to-event outcomes, alongside with its marginalized version (Chapter 5).

Nowadays, many research is done on joint modeling, due to the large amount of possibilities. While this thesis places focus on four cases, its convenient to say that lot of research still needs to be done on this topic. For example, due to the flexibility of the WGN model, a research gap opens to the incorporation of informative censoring (Section 6.5.2), an issue that was problematic in the past when modeling time-to-event data. Moreover, considerable research is still ongoing on exploring goodness-of-fit tests and diagnostic tools for the proposed WGN model. For example, Drivandi, Verbeke and Molenberghs (2014) recently developed a novel diagnostic test based on the gradient function of Section 8.2 to assess the random-effects distribution. They established asymptotic properties for their test and showed that the proposed test statistic converges to a weighted sum of independent chi-squared random variables each with one degree of freedom under a correctly specified

random-effects distribution. Additionally, they developed a parametric bootstrap algorithm for small samples, and denoted that the strategy can be used to check the adequacy of any distribution for random effects in a wide class of mixed moels, including LMM, GLMM and non-linear mixed models, with univariate as well as multivariate random effects.

However, as has been pointed out in Chapter 7 & 10, sensitivity analysis seems to be a recurrent theme. Due to the strong connection between the longitudinal and time-to-event setting, and the missing data one, an extended shared random effects joint model (Njagi et al, 2013c) is explored in Chapter 7, where a characterization of MAR is provided within. While the framework has been built conceptually, the elaborate random-effects structure has been used as an avenue for sensitivity analysis in this context. Given the interrelationships that arise among model components, model formulation under the framework studied in Chapter 7 becomes exceedingly complex. The effect of such difficulties on the studied framework, especially on the results established, and the possibility of expanding the framework to take into account such difficulties, requires further investigation.

The analysis made for the asthma dataset showed that the extended framework increased in model fit, when comparing it to traditional modeling frameworks. Therefore, its important to note that aspects like overdispersion and hierarchical structure need to be taken into account when making appropriate predictions and conclusions. Since general conclusions cannot be made on a few data analysis, simulation studies are put forward to explore the extended framework in detail. To conclude, it should be said that once should not consider the framework as best fit, but more as an elegant way of dealing with overdispersion and hierarchical structure simultaneously.

# References


[1] Agresti A. *Categorical data analysis (2nd ed.)*. John Wiley & Sons, Inc, New York, 2002.

[2] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G., and Kenward M. G. Sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125, 2010.

[3] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G., and Kenward M. G. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125, 2010.

[4] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G., and Kenward M. G. Generalized shared-parameter models and missingness at random. *Statistical Modelling*, 11(4):279–310, 2011.

[5] Efendi A. and Molenberghs G. A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *Journal of Biopharmaceutical Statistics*, 13(6):1420–1434, 2013.

[6] Efendi A. and Molenberghs G. A multilevel model for hierarchical, repeated, overdispersed time-to-event outcomes and its estimation strategies. *Journal of Biopharmaceutical Statistics*, 23(6):1420–1434, 2013.

[7] Efendi A., Molenberghs G., Njagi E. N., and Dendale P. A joint model with marginal interpretation for longitudinal continuous and time-to-event outcomes. *Statistical Modelling*, 2013.

[8] Efendi A., Molenberghs G., and Iddi S. A marginalized combined model for repeated, overdispersed time-to-event outcomes. *Communication in Statistics - Theory and Methods*, 43(22):4806–4828, 2014.

[9] Efendi A., Verbeke G., and Molenberghs G. A goodness-of-fit test for the random-effects distribution in mixed models. *Statistical Methods in Medical Research*, 2013.

[10] Gelman A. and Rubin D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–511, 1992.

[11] Gelman A., Carlin J. B., Stern H. S., and Rubin D. B. *Bayesian data analysis*. Texts in Statistical Science, Chapman & Hall, London, 1995.

[12] Ivanova A., Molenberghs G., and Verbeke G. A model for overdispersed hierarchical ordinal data. *Statistical Modelling*, 14(5):399–415, 2014.

[13] Skrondal A. and Rabe-Hesketh S. *Generalized latent variable modeling*. Chapman & Hall/CRC, London, 2004.

[14] Tsiatis A. and Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(1):809–834, 2004.

[15] Zuur A., Leno E. N., and Smith G. M. *Analysing ecological data*. Spinger, New York, 2007.

[16] Zuur A., Leno E. N., Smith G. M., Saveliev A. A., and Walker N. *Mixed effects models and extensions in ecology with R*. Spinger, New York, 2009.

[17] Lawson A. B. and Browne W. J. *Disease mapping with WinBUGS and MLwiN (1st ed.) (Statistics in Practice)*. John Wiley & Sons, Inc, New York, 2003.

[18] Gelfand A. E. and Smith A. F. M. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

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

[20] Ghebretinsae A. H., Faes C., Molenberghs G., De Boeck M., and Geys H. A bayesian, generalized frailty model for comet assays. *Journal of Biopharmaceutical Statistics*, 23(3):618–636, 2013.

[21] Lemonte A. J. and A. G. Patriota. Multivariate elliptical models with general parameterization. *Statistical Methodology*, 8(4):389–400, 2011.

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

[23] J$\phi$rgensen B. Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B*, 49(2):127–162, 1987.

[24] Arnold B. C. and Strauss D. Pseudolikelihood estimation: Some examples. *The Indian Journal of Statistics, Series B (1960-2002)*, 53(2):233–243, 1991.

[25] Lindsay B. G. Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–239, 1988.

[26] Spearman C. "general intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904.

[27] Varin C. On composite marginal likelihoods. *Advances in Statistical Analysis*, 92(1):233–243, 2008.

[28] Dean C. B. Estimating equations for mixed-poisson models. *Estimating Functions, V.P. Godambe (Ed.)*, 1(1):35–46, 1991.

[29] Dean C. B. and Nielsen J. D. Generalized linear mixed models: A review and some extensions. *Lifetime Data Analysis*, 13(4):497–512, 2007.

[30] Geyer C. J. Practical markov chain monte carlo. *Statistical Science*, 7(4):473–483, 1992.

[31] Wintrebert C. M. A., Zwinderman A. H., Cam E., Pradel R., and van Houwelingen J. C. Joint modelling of breeding and survival in the kittiwake using frailty models. *Ecological Modelling*, 181(2-3):203–213, 2005.

[32] Clayton D. and Rasbash J. Estimation in large crossed random-effect models by data augmentation. *Journal of the Royal Statistical Society, Series A*, 162(3):425–436, 1999.

[33] Cox D. and Reid N. A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737, 2004.

[34] Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.

[35] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

[36] Rizopoulos D. *Joint models for longitudinal and time-to-event data*. Boca Raton: Chapman and Hall/CRC, 2012a.

[37] Rizopoulos D., Verbeke G., and Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74, 2008.

[38] Stegmueller D. How many countries for multilevel modeling? a comparison of frequentist and bayesian approaches. *American Journal of Political Science*, 57(3):748–761, 2013.

[39] Hall D. B. Zero-inflated poisson and binomial regression with random effects: A case study. *Biometrics*, 56(4):1030–1039, 2000.

[40] Rubin D. B. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[41] Rubin D. C., Hinton S., and Wenzel A. The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5):1161–1176, 1999.

[42] Heitjan D. F. and Rubin D. B. Ignorability and coarse data. *The Annals of Statistics*, 19(4):2244–2253, 1991.

[43] Cox D. R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34(2):187–220, 1972.

[44] Cox D. R. Some remarks on overdispersion. *Biometrika*, 70(1):269–274, 1983.

[45] Cox D. R. and Hinkley D. V. *Theoretical statistics*. Chapman & Hall, London, 1974.

[46] Lesaffre E. and Lawson A. B. *Bayesian biostatistics*. John Wiley & Sons, Ltd, New York, 2012.

[47] Lesaffre E. and Verbeke G. Local influence in linear mixed models. *Biometrics*, 54(2):570–582, 1998.

[48] Norden E. H., Long S. R., Tung C. C., Yuan Y., and Bliven L. F. A non-gaussian statistical model for surface elevation of nonlinear random wave fields. *Journal of Geophysical Research*, 88(C12):7597–7606, 1983.

[49] Njagi E. N., Rizopoulos D., Molenberghs G., Dendale P., and Willekens K. A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modelling*, 13(3):179–198, 2013.

[50] Njagi E. N., Molenberghs G., Rizopoulus D., Verbeke G., Kenward M. G., Dendale P., and Willekens K. A flexible joint-modeling framework for longitudinal and time-to-event data with overdispersion. *Statistical Methods in Medical Research*, 0(0):1–16, 2013.

[51] Njagi E. N., Molenberghs G., Kenward M. G., Verbeke G., and Rizopoulus D. A characterization of missingness at random in a generalized shared-parameter joint modelling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, 56(6):1001–1015, 2013.

[52] Tuerlinckx F., Rijmen F., Molenberghs G., Verbeke G., Briggs D., Van Den Noortgate W., Meulders M., and De Boeck P. Estimation and software. in p. de boeck & m. wilson (eds.). *Explanatory item response models: A generalized linear and nonlinear approach*, 12(1):343–373, 2004.

[53] Hildebrand F. B. *Advanced calculus for applications.* Prentice-Hall, Englewood Cliffs, 1962.

[54] D'Agostini G. Bayesian inference in processing experimental data: Principles and basic applications. *Reports on Progress in Physics*, 66(9):1383–1419, 2003.

[55] Frahm G. Generalized elliptical distributions: Theory and applications. *Thesis*, 2004.

[56] Molenberghs G., Michiels B., Kenward M. G., and Diggle P. J. Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161, 1998.

[57] Molenberghs G., Beunckens C., Sotto C., and Kenward M. G. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, 70(2):371–388, 2008.

[58] Molenberghs G. and Lesaffre E. Marginal modelling of correlated ordinal data using an n-way plackett distribution. *Journal of the American Statistical Association*, 78(1):139–144, 1994.

[59] Molenberghs G., Njagi E. N., Kenward M. G., and Verbeke G. Enriched-data problems and essential non-identifiability. *International Journal of Statistics in Medical Research*, 1(1):16–44, 2012.

[60] Molenberghs G. and Verbeke G. *Models for discrete longitudinal data*. Springer, New York, 2005.

[61] Molenberghs G., Verbeke G., Efendi A., Braekers R., and Demétrio C. G. B. A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *Statistical Methods in Medical Research*, 0(0):1–19, 2014.

[62] Molenberghs G., Verbeke G., and Demétrio C. G. B. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, 13(4):513–531, 2007.

[63] Molenberghs G., Verbeke G., Demétrio C. G. B., and Vieira A. M. C. A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, 25(3):325–347, 2010.

[64] Molenberghs G. and Kenward M. G. *Missing data in clinical studies*. John Wiley & Sons, Inc, 2007.

[65] Smith G. *Statistical reasoning (3rd ed.)*. Allyn and Bacon, Boston, 1991.

[66] Verbeke G. and Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221, 1996.

[67] Verbeke G. and Molenberghs G. *Linear mixed models for longitudinal data*. Springer, New York, 2000.

[68] Verbeke G. and Molenberghs G. Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling*, 10(4):391–419, 2010.

[69] Verbeke G. and Molenberghs G. The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics*, 1(14):477–490, 2013.

[70] Verbeke G., Molenberghs G., and Rizopoulos D. Random effects models for longitudinal data. *Longitudinal Research with Latent Variables, van Montfort, K., Oud, H., & Satorra, A.(Eds.)*, pages 37–96 (Chapter 2), 2004.

[71] Verbeke G., Molenberghs G., Thijs H., Lesaffre E., and Kenward M. G. Sensitivity analysis for non-random dropout: A local influence approach. *Biometrics*, 57(1):7–14, 2001.

[72] Schwarz G. E. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[73] Box G. E. P. and Tiao G. C. *Bayesian inference in statistical analysis*. John Wiley & Sons, Inc, New York, 1992.

[74] Fitzmaurice G. M., Davidian M., Molenberghs G., and Verbeke G. *Longitudinal Data Analysis*. Handbooks of Modern Statistical Methods, Chapman & Hall/CRC, New York, 2009.

[75] M. Galea, G. A. Paula, and H. Bolfarine. Local influence in elliptical linear regression models. *The Statistician*, 46(1):71–79, 1997.

[76] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[77] Geys H., Molenberghs G., and Lipsitz S. R. A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *Journal of Statistical Computation and Simulation*, 62(1–2):45–72, 1998.

[78] Joe H. and Lee Y. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100(4):670–685, 2008.

[79] Tao H., Palta M., Yandell B. S., and Newton M. A. An estimation method for the semiparametric mixed effects model. *Biometrics*, 55(1):102–110, 1999.

[80] Van Oyen H. and Tafforeau J. Health interview surveys. *Arch Public Health*, 52(1):79–82, 1994.

[81] Van Oyen H., Tafforeau J., Hermans H., Quataert P., E. Schiettecatte, Lebrun L., and Belamer L. The belgian health interview survey. *Arch Public Health*, 55(1):1–13, 1997.

[82] Zhu H. T. and Lee S. Y. Local influence for incomplete-data models. *Journal of the Royal Statistical Society. Series B*, 63(1):111–126, 2001.

[83] Hammami I., Garcia A., and Nuel G. Evidence for overdispersion in the distribution of malaria parasites and leukocytes in thick blood smears. *Malaria Journal*, 12(398), 2013.

[84] Myung I. J. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100, 2002.

[85] Berger J. The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385–402, 2006.

[86] Chen J., Zhang D., and Davidian M. A monte carlo em algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, 3(3):347–360, 2002.

[87] Engel J. Models for response data showing extra-poisson variation. *Statistica Neerlandica*, 38(3):159–167, 1984.

[88] Hinde J. and Demétrio C. G. B. Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27(2):151–170, 1998a.

[89] Hinde J. and Demétrio C. G. B. *Overdispersion: Models and estimation.* XIII Sinape, São Paulo, 1998b.

[90] Rasbash J., Steele F., Browne W. J., and Goldstein H. *A user's guide to MLwiN.* Centre for Multilevel Modelling, University of Bristol, Bristol, 2014.

[91] Zhang J. and Heitjan D. F. Impact of nonignorable coarsening on bayesian inference. *Biostatistics*, 8(4)(1):722–743, 2007.

[92] Nelder J. A. and Wedderburn R. W. M. Generalized linear models. *Journal of the Royal Statistical Society (Series A)*, 135(3):370–384, 1972.

[93] Pinheiro J. C. and Bates D. M. *Mixed-effects models in S and S-PLUS*. Statistics and Computing Series, Springer-Verlag, New York, 2000.

[94] Jansen J. F., Stambuk H. E., Koutcher J. A., and Shukla-Dave A. Non-gaussian analysis of diffusion-weighted mr imaging in head and neck squamous cell carcinoma: A feasibility study. *American Journal of Neuroradiology*, 31(4):741–748, 2010.

[95] Lawless J. F. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, 15(3):209–225, 1987.

[96] Booth J. G., Casella G., Friedl H., and Hobert J. P. Negative binomial loglinear mixed models. *Statistical Modelling*, 3(3):179–181, 2003.

[97] Skellam J. G. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):257–261, 1948.

[98] Faraway J. J. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall/CRC, Boca Raton, 2006.

[99] Lambrechts K. Information-accumulation theory of speeded categorization. *Psychological Review*, 107(2):227–260, 2000.

[100] Van Steen K., Molenberghs G., Verbeke G., and Thijs H. A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling*, 1(2)(1):125–142, 2001.

[101] Manton K. G., Woodbury M. A., and Stallard E. A variance components approach to categorical data models with heterogenous cell populations: Analysis of spatial gradients in lung cancer mortality rates in north carolina counties. *Biometrics*, 37(2):259–269, 1981.

[102] Liang K. Y. and Zeger S. L. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[103] Duchateau L. and Jassen P. *The frailty model*. Springer, New York, 2007.

[104] Magder L. S. and Zeger S. L. A smooth nonparametric estimate of a mixing distribution using mixtures of gaussians. *Journal of the American Statistical Association*, 91(435):1141–1151, 1996.

[105] Aerts M., Claeskens G., Hens N., and Molenberghs G. Local multiple imputation. *Biometrika*, 89(2):375–388, 2002.

[106] Aregay M., Shkedy Z., and Molenberghs G. A hierarchical bayesian overdispersion models for non-gaussian repeated measurement data. *Computational Statistics and Data Analysis, Submitted*, 2013.

[107] Aregay M., Shkedy Z., and Molenberghs G. A hierarchical bayesian approach for analysis of longitudinal count data with overdispersion parameters: A simulation study. *Computational Statistics and Data Analysis*, 57(1):233–245, 2013a.

[108] Davidian M. and Gallant A. R. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80(3):475–488, 1993.

[109] De Backer M., De Keyser P., De Vroey C., and E. Lesaffre. A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/daya double-blind comparative trial. *British Journal of Dermatology*, 134(46)(38):16–17, 1996.

[110] Evans M. and Swartz T. Methods for approximating integrals in statistics with special emphasis on bayesian integration problems. *Statistical Science*, 10(3):254–272, 1995.

[111] Haker M., Bohme M., Martinetz T., and Barth E. Geometric invariants for facial feature tracking with 3d tof cameras. *International Symposium on Signals, Circuits and Systems*, 1(1):1–4, 2007.

[112] Griswold M. E. and Zeger S. L. On marginalized multilevel models and their computation. *John Hopkins Bloomberg School of Public Health*, 99(1):319–331, 2004.

[113] Kenward M. G., Goetghebeur E. J. T., and Molenberghs G. Sensitivity analysis for incomplete categorical data. *Statistical Modelling*, 1(1):31–48, 2001.

[114] Kenward M. G., Goetghebeur E. J. T., and Molenberghs G. Sensitivity analysis for incomplete categorical data. *Statistical modelling*, 1(1):31–48, 2001.

[115] Ouwens M. J. N. M., Tan F. E. S., and Berger M. P. F. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, 57(4):1166–1172, 2001.

[116] Breslow N. E. Extra-poisson variation in log-linear models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 33(1):38–44, 1984.

[117] Breslow N. E. Tests of hypotheses in overdispersed poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association*, 85(410):565–571, 1990.

[118] Breslow N. E. and Clayton D.G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(412):9–25, 1993.

[119] Breslow N. E. and Lin X. Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, 82(1):81–91, 1995.

[120] Laird N. M. and Ware J. H. Random effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[121] McCullagh P. and Nelder J. *Generalized linear models (2nd ed.)*. Chapman & Hall/CRC, Boca Raton, 1989.

[122] Thall P. F. and Vail S. C. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46(3):657–671, 1990.

[123] Davis P. J. and Rabinowitz P. *Methods of numerical integration (2nd ed.)*. Academic Press, Dover, 1984.

[124] Diggle P. J., Heagerty P. J., Liang K. Y., and Zeger S. L. *Analysis of Longitudinal Data*. Oxford University Press, Oxford, 2002.

[125] Heagerty P. J. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, 55(3):688–698, 1999.

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

[127] Christensen R. Principals of statistical inference. *Journal of the American Statistical Association*, 103(484):1719–1723, 2008.

[128] Ecochard R. and Clayton D. J. Fitting complex random effect models with standard software using data augmentation: Application to a study of male and female fecundability. *Statistical Modelling,*, 1(4):319–331, 2002.

[129] Henderson R., Diggle P., and Dobson A. Joint modeling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

[130] Schall R. Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–727, 1991.

[131] Swinburne R. *Bayess theorem*. Oxford University Press, Oxford, 2002.

[132] Willink R. Normal moments and hermite polynomials. *Statistics and Probability Letters*, 73(3):271–275, 2005.

[133] Wolfinger R. and O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48(3):233–243, 1993.

[134] Littell R. C., Stroup G. A., Wolfinger W. W., and Schabenberger O. *SAS for mixed models (2nd ed.)*. SAS Institute Inc, Cary, 2006.

[135] Cook R. D. Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, 48(2):133–169, 1986.

[136] Wolfinger R. D. and Chang M. *Comparing the SAS GLM and MIXED procedures for repeated measures*. SAS Institute Inc, Boca Raton, 1998.

[137] Wolfinger R. D. and Lin X. Two taylor-series approximation methods for nonlinear mixed models. *Computational Statistics & Data Analysis archive*, 25(4):465–490, 1997.

[138] Bahadur R. R. A representation of the joint distribution of responses to $n$ dichotomous items. *Studies in item analysis and prediction*, 1(1):158–168, 1961.

[139] Wilcox R. R. A review of the beta-binomial model and its extensions. *Journal of Educational Statistics*, 6(1):3–32, 1981.

[140] Wedderburn R. W. M. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447, 1974.

[141] Chib S. and Greenberg E. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[142] Fieuws S. and Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine*, 23(20):3093–3104, 2004.

[143] Geman S. and Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[144] Iddi S. and Molenberghs G. A combined overdispersed and marginalized multilevel model. *Computational Statistics and Data Analysis*, 56(6):1944–1951, 2012.

[145] Iddi S. and Molenberghs G. A joint marginalized multilevel model for longitudinal outcomes. *27th International Workshop on Statistical Modelling (Vol. 1)*, 39(11):165–170, 2012.

[146] Iddi S., Molenberghs G., Aregay M., and Kalema G. Empirical bayes estimates for correlated hierarchical data with overdispersion. *Pharmaceutical Statistics*, 13(5):316–326, 2014.

[147] le Cessie S. and van Houwelingen J. C. A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, 47(4):1267–1282, 1991.

[148] Liu S. On local influence in elliptical linear regression models. *Statistical Papers*, 41(1):211–224, 2000.

[149] Liu S. Local influence in multivariate elliptical linear regression models. *Linear Algebra and its Applications*, 354(1–3):159–174, 2002.

[150] Cho S. J. and Rabe-Hesketh S. Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics and Data Analysis*, 55(1):12–25, 2011.

[151] Zeger S. L. and Liang K. Y. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, 11(14–15):1825–1839, 1992.

[152] Zeger S. L., Liang K. Y., and Albert P. S. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44(4):1049–1060, 1988.

[153] Stigler S. M. The epic story of maximum likelihood. *Statistical Science*, 22(4):598–620, 2007.

[154] Paul S. R., Plackett R. L., and Wedderburn R. W. M. Inference sensitivity for poisson mixtures. *Biometrika*, 65(3):591–602, 1978.

[155] Raudenbush S. W. Marginalized multilevel models and likelihood inference: Comments. *Statistical Science*, 15(1):22–24, 2000.

[156] Raudenbush S. W., Yang M. L., and Yosef M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, 9(1):141–157, 2000.

[157] Rakhmawati T., Molenberghs G., Verbeke G., and C. Faes. Local influence diagnostics for generalized linear mixed models with overdispersion. *Statistical Modelling*, 2014.

[158] Mwangi T. W., Fegan G., Williams T. N., Kinyanjui S.M., Snow R. W., and Marsh K. Evidence for over-dispersion in the distribution of clinical malaria episodes in children. *Plos One*, 3(5):21–96, 2014.

[159] Rakhmawati T. W., Molenberghs G., Verbeke G., and Faes C. Local influence diagnostics for generalized linear mixed models with overdispersion. *Not been published yet*, 2014.

[160] Gardner W., William M., Mulvey E. P., and Shaw E. C. Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, 118(3):392–404, 1995.

[161] Shih W. J. Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med.*, 3(1)(4), 2002.

[162] Hastings W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[163] Lin X. Variance component testing in generalised linear models with random effects. *Biometrika*, 84(2):309–326, 1997.

[164] Lee Y., Nelder J. A., and Pawitan Y. *Generalized linear models with random effects: Unified analysis via H-likelihood*. Chapman & Hall/CRC, Boca Raton, 2006.

[165] Tseng Y. K., Hsieh F., and Wang J. L. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92(3):587–603, 2005.

# Chapter A

# Supplementary Material for Chapter 3

## A.1 Derivation of the Mean and Variance for the Exponential Family

**Theorem A.1.1.** *Suppose a random variable $Y$ that follows the exponential family with density*

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

*Then, the mean and variance of $Y$ equals*

$$E(Y) = \psi'(\eta), \tag{A.2}$$

$$Var(Y) = \phi \cdot \psi''(\eta), \tag{A.3}$$

*Proof.* (for the continuous case) The random variable $Y$ possesses the characteristic that the area under the density function equals 1:

$$\int_{-\infty}^{+\infty} f(y \mid \eta, \phi)dy = \int_{-\infty}^{+\infty} \exp\{\phi^{-1} \cdot [y \cdot \eta - \psi(\eta)] + c(y, \phi)\}dy = 1.$$

Calculating the first and second derivates of this integral w.r.t. $\eta$ gives:

$$\begin{cases} \frac{\partial}{\partial \eta} \int_{-\infty}^{+\infty} f(y \mid \eta, \phi)dy = 0 \\ \frac{\partial^2}{\partial \eta^2} \int_{-\infty}^{+\infty} f(y \mid \eta, \phi)dy = 0 \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \int_{-\infty}^{+\infty} [y - \psi'(\eta)] \cdot f(y \mid \eta, \phi)dy = 0 \\ \int_{-\infty}^{+\infty} [\phi^{-1} \cdot (y - \psi'(\eta))^2 - \psi''(\eta)] \cdot f(y \mid \eta, \phi)dy = 0 \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \mathrm{E}(Y) = \psi'(\eta) \\ \mathrm{Var(Y)} = \phi \cdot \psi''(\eta) \end{cases}$$

Trivial derivations can be done for the discrete case, where the integral is replaced by the summation. ■

## A.2 Generic Approximations for Marginal Model Elements in the Combined Model Framework

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

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

i.e.,

$$\mathrm{E}(Y_{ij}) \approx \varpi_{ij} \cdot \left[ h(\eta_{ij}^{(0)}) + \frac{1}{2} \cdot h''(\eta_{ij}^{(0)}) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij} \right], \tag{A.4}$$

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

$$\mathrm{Var}(Y_{ij}) = \mathrm{E}\{\mathrm{E}[\mathrm{Var}(Y_{ij} \mid \mathbf{b}_i, \theta_{ij})]\} + \mathrm{E}\{\mathrm{Var}[\mathrm{E}(Y_{ij} \mid \mathbf{b}_i, \theta_{ij})]\} + \mathrm{Var}\{\mathrm{E}[\mathrm{E}(Y_{ij} \mid \mathbf{b}_i, \theta_{ij})]\},$$

To simplify ensuing derivations, write the variance function as

$$\omega(\mu_{ij}^c) = \omega_{ij}(\theta_{ij} \cdot \kappa_{ij}) = \phi \cdot \psi''[g(\theta_{ij} \cdot \kappa_{ij})],$$

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

$$\begin{aligned}
\mathrm{Var}(Y_{ij}) \approx\ & \omega\left[h(\eta_{ij}^{(0)})\right] + \omega'\left[h(\eta_{ij}^{(0)})\right] \cdot (\varpi_{ij} - 1) + \frac{1}{2} \cdot \omega''\left[h(\eta_{ij}^{(0)})\right] \cdot h^2(\eta_{ij}^{(0)}) \cdot (\sigma_{ij}^2 + \varpi_{ij}^2 \\
& - 2 \cdot \varpi_{ij} + 1) + \frac{1}{2} \cdot \omega''\left[h(\eta_{ij}^{(0)})\right] \cdot h(\eta_{ij}^{(0)}) \cdot h'(\eta_{ij}^{(0)}) \cdot (\varpi_{ij}^2 + \sigma_{ij}^2) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij} \\
& + \sigma_{ij}^2 \cdot h^2(\eta_{ij}^{(0)}) + \frac{1}{2} \cdot \omega'\left[h(\eta_{ij}^{(0)})\right] \cdot h''(\eta_{ij}^{(0)}) \cdot \varpi_{ij} \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij} + \sigma_{ij}^2 \cdot \left[h'(\eta_{ij}^{(0)})\right]^2 \\
& \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij} + \sigma_{ij}^2 \cdot \left[h'^2(\eta_{ij}^{(0)}) + h(\eta_{ij}^{(0)}) \cdot h''(\eta_{ij}^{(0)})\right] \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij}.
\end{aligned} \tag{A.5}$$

Likewise, for the covariance function:

$$\begin{aligned}
\mathrm{Cov}(Y_{ij}, Y_{ik}) \approx\ & \sigma_{ijk} \cdot \left[ h(\eta_{ij}^{(0)}) \cdot h(\eta_{ik}^{(0)}) + \frac{1}{2} \cdot h(\eta_{ij}^{(0)}) \cdot h''(\eta_{ik}^{(0)}) \cdot \mathbf{z}_{ik}^{'} \cdot D \cdot \mathbf{z}_{ik} + \frac{1}{2} \cdot h(\eta_{ik}^{(0)}) \right. \\
& \left. \cdot h''(\eta_{ij}^{(0)}) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij} + h'(\eta_{ij}^{(0)}) \cdot h'(\eta_{ik}^{(0)}) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ik} \right] + \varpi_{ij} \cdot \sigma_{ik}^2 \cdot h'(\eta_{ij}^{(0)}) \\
& \cdot h'(\eta_{ik}^{(0)}) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ik}.
\end{aligned} \tag{A.6}$$

Here, $\sigma_{ijk}$ is the covariance between $\theta_{ij}$ and $\theta_{ik}$. In case these effects are assumed to be independent, a large portion of $(A.6)$ then cancels, with covariance induced solely by the effects $\mathbf{b}_i$. In case all $\theta_{ij}$ are equal, $\sigma_{ijk} \equiv \sigma_{ij}$. Evidently, $(A.5)$ and $(A.6)$ lead to approximate expressions for the correlations, too. Of course, in situations where closed forms exist, these expressions need not be used.

Needless to say that the above approximations may or may not be accurate, depending on the context. Therefore, their use should be seen as poor man's choice, when no explicit forms are available. Fortunately, closed forms are available for the Weibull (Section 4.1) case.

# Chapter B

# Supplementary Material for Chapter 4

## B.1 Marginal Density and Moments of the Weibull-Gamma-Normal Model

In this part of the thesis, the marginal density of the model specified by (4.1)–(4.3) is derived. First replace the predictor $\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i$ in (4.1) by $\mu$, and integrate

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

over the general gamma distribution of $\theta$, i.e., over one component of (4.2):

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

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

Now, for the general case, first observe that

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

It then follows that

$$\begin{aligned} f(\mathbf{y}_i \mid \boldsymbol{\theta}_i) = & \prod_{j=1}^{n_i} \lambda \cdot \rho \cdot \theta_{ij} \cdot y_{ij}^{\rho-1} \cdot \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \\ & \cdot \int_{b} e^{\mu_{ij} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i} \cdot e^{-\lambda \cdot y_{ij}^{\rho} \cdot \theta_{ij} \cdot e^{\mu_{ij} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i}} \cdot e^{\frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i} \cdot d\mathbf{b}_i \qquad \text{(B.2)} \\ = & \sum_{(m_1, \ldots, m_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1) \cdot \rho - 1} \cdot \theta_{ij}^{m_j+1} \\ & \cdot \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot \int_{b} e^{(m_j+1) \cdot (\mu_{ij} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i) - \frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i} \cdot d\mathbf{b}_i. \qquad \text{(B.3)} \end{aligned}$$

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

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

with

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

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

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

Further integration over the gamma distribution produces

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

with

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

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

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

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

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

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

As for the moments, the moments are first derived based upon $(B.1)$:

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

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

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

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

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

where rewriting

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

with

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

From $(B.12)$, the following moment expression, with mean, variance, and covariance expressions are immediately derived:

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

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

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

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

It is customary, in the standard frailty model (Duchateau and Janssen, 2007), to set $\alpha_j \cdot \beta_j = 1$, for identifiability purposes (also suggested in Section 3.2). The change to $(B.8)$ on the one hand, and to $(B.13)$–$(B.16)$ is then both evident and minor. Likewise, the exponential version follows from setting $\rho = 1$. While Molenberghs and Verbeke (2011) showed that only a finite number of moments is finite, it is wise to check the number of finite moments for the WGN model. Using formula $(B.13)$ for the WGN moments, Molenberghs et al (2014) denoted that the order $k \leq \alpha_j \cdot \rho$ for the corresponding moment to be finite.

## B.2 Regularity Conditions, Consistency and Asymptotic Normality of the Pseudo-likelihood Estimator

The logarithm of the pseudo-likelihood (4.11) can be maximized if the density functions $f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi})$ obtain the following six regularity conditions:

**B1** The densities $f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi})$ are distinct for different values of the parameter $\boldsymbol{\xi}$.

**B2** The densities $f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi})$ have common support, which does not depend on $\boldsymbol{\xi}$.

**B3** The parameter space $\Omega$ contains an open region $\omega$ of which the true parameter value $\boldsymbol{\xi}_0$ is an interior point.

**B4** $\omega$ is such that for all $s$, and almost all $\mathbf{y}^{(s)}$ in the support of $\mathbf{Y}^{(s)}$, the densities admit all third derivatives, for $j, k, l = 1, \ldots, p$,

$$\frac{\partial^3 f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi})}{\partial \xi_j \partial \xi_k \partial \xi_l}. \tag{B.17}$$

**B5** The first and second logarithmic derivatives of $f_s$ satisfy, for $k, l = 1, \ldots, p$,

$$\mathrm{E}_{\boldsymbol{\xi}} \left\{ \frac{\partial \ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]}{\partial \xi_k} \right\} = 0, \qquad \text{and} \qquad 0 < \mathrm{E}_{\boldsymbol{\xi}} \left\{ \frac{-\partial^2 \ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]}{\partial \xi_k \partial \xi_l} \right\} < +\infty. \tag{B.18}$$

**B6** The matrix $I_0$, defined in $(B.20)$, is positive definite.

**B7** There exist functions $M_{klr}$ such that

$$\sum_{s \in S} \delta_s \cdot \mathrm{E}_{\boldsymbol{\xi}} \left| \frac{\partial^3 \ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]}{\partial \xi_k \partial \xi_l \partial \xi_r} \right| < M_{klr}(\mathbf{y}) \tag{B.19}$$

for all $\mathbf{y}$ in the support of $f$ and for all $\theta \in \omega$ and $m_{klr} = \mathrm{E}_{\boldsymbol{\xi}_0}(M_{klr}(Y)) < +\infty$.

Theorem B.2.1, proven by Arnold and Strauss (1991), guarantees the existence of at least one solution to the pseudo-likelihood equations, which is a consistent and asymptotically normal estimator. Without loss of generality, assume $\boldsymbol{\xi}$ is constant. Replacing it by $\boldsymbol{\xi}_i$, and modeling it as a function of covariates is straightforward.

**Theorem B.2.1.** (Consistency and Asymptotic Normality) *Assume that* $(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$ *are i.i.d. with common density that depends on* $\boldsymbol{\xi}_0$. *Then under regularity conditions* (**B1**)–(**B7**)*:*

*[1] the pseudo-likelihood estimator* $\tilde{\boldsymbol{\xi}}_N$, *defined as the maximizer of* (4.11), *converges in probability to* $\boldsymbol{\xi}_0$.

*[2]* $\sqrt{N} \cdot (\tilde{\boldsymbol{\xi}}_N - \boldsymbol{\xi}_0)$ *converges in distribution to* $N_p(\mathbf{0}, I_0(\boldsymbol{\xi}_0)^{-1} \cdot I_1(\boldsymbol{\xi}_0) \cdot I_0(\boldsymbol{\xi}_0)^{-1})$ *with* $I_0(\boldsymbol{\xi})$ *defined by*

$$I_{0,kl}(\boldsymbol{\xi}) = -\sum_{s \in S} \delta_s \cdot E_{\boldsymbol{\xi}} \left\{ \frac{\partial^2 ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]}{\partial \xi_k \partial \xi_l} \right\} \tag{B.20}$$

*and* $I_1(\boldsymbol{\xi})$ *by*

$$I_{1,kl}(\boldsymbol{\xi}) = \sum_{s,t \in S} \delta_s \cdot \delta_t \cdot E_{\boldsymbol{\xi}} \left\{ \frac{\partial ln \left[ f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\xi}) \right]}{\partial \xi_k} \cdot \frac{\partial ln \left[ f_t(\mathbf{y}_i^{(t)}; \boldsymbol{\xi}) \right]}{\partial \xi_l} \right\}. \tag{B.21}$$

## B.3 SAS Macro for the Two-Stage Approach

For providing the SAS macro of Iddi et al (2014), i.e., {*proc iml; ...; quit;*}, in a clear way, the Asthma dataset (Duchateau and Janssen, 2008) will be used and analyzed here. The time to recurrence of an asthma attack, denoted by $Time_{ij}$, is modeled with the following (random-intercept) WGN model:

$$Time_{ij} \mid \mathbf{b}_i, \theta_{ij} \sim \text{Weibull}(\rho, k_{ij}), \tag{B.22}$$

$$k_{ij} = \theta_{ij} \cdot e^{\xi_0 + \xi_1 \cdot T_i + b_i}, \tag{B.23}$$

$$b_i \sim N(0, \sigma^2), \tag{B.24}$$

$$\theta_{ij} \sim \text{Gamma}(\alpha, 1/\alpha), \tag{B.25}$$

where $T_i = 0$ if patient $i$ got a placebo and 1 if patient $i$ got the drug. The corresponding SAS code is given below:

```
/*************************************************************
SOFTWARE: SAS 9.3.
OBJECTIVE: Providing the SAS macro of Iddi et al (2014) to obtain
EB Gamma frailty estimates for analyzing the Asthma dataset using
the WGN model (4.1)-(4.3), assuming full likelihood with incorpo-
rating censoring;
DATASET: Example 9 of Duchateau & Janssen (2008);
VARIABLE DESCRIPTION:
Patid: Patient ID;
```

```
Begin and End: time interval between events for each patient;
Status: Right censoring indicator (1 = Asthma Attack, 0 = Censo-
red);
Drug: Treatment indicator (1 = Drug, 0 = Placebo).
REFERENCES: Iddi et al (2014) & Molenberghs et al (2013).

Thanks given to Samuel Iddi & Achmad Efendi, for providing rele-
vant SAS code.
*************************************************************/
libname asthma 'D:\Mijn Documenten\Master Statistiek\Masterthesis
               \Data';
libname u 'D:\Mijn Documenten\Master Statistiek\Masterthesis\EB
           schatting';

/*Weibull-Gamma-Normal model - full likelihood with right censo-
ring*/
proc nlmixed data = asthma.asthma1 tech = quanew qpoints = 50
             maxit = 1000;
        bounds lambda > 0, alpha > 0;
        parms Beta_0 = -2 Beta_1 = -0.16 lambda = 1 rho = 1
              alpha = 3.3 sigma = 1;
        eta = (Beta_0 + b) + Beta_1*(Drug=1);
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha) + (rho
             -1)*log(Time) + eta - (alpha+1)*log(lambda*(Time**
             rho)*expeta + alpha);
        loglik  = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(loglik);
        random b ~ normal(0, sigma**2) subject=Patid out=u.biwei;
        estimate 'Variance of R.E.s' sigma**2;
run;

/*Integration over theta_ij (Iddi et al, 2014)*/
proc iml;
        use asthma.asthma1;
        read all into data;
        M = nrow(data);

        start fxn1(theta) global(alpha, lambda, rho, subdata, i,
                          j, b0, bi);
              Drug = subdata[, 5];
              Time = subdata[, 7];
              term1 = b0[1] + b0[2]#Drug[j] + bi[i];
              kappa = theta#exp(term1);
              fy = lambda#rho#kappa#Time[j]##(rho-1)#exp(-
                   lambda#Time[j]##rho#kappa);
```

```
        term = (1/alpha)##alpha#exp(lgamma(alpha));
        ftheta = (1/term)#theta##(alpha-1)#exp(-theta
                 /(1/alpha));
        int = fy#ftheta;
return(int);
finish;

start fxn2(theta) global(alpha, lambda, rho, subdata, i,
                  j, b0, bi);
        num = fxn1(theta);
        eps = 1E-10;
        lim2 = {0 .P};
        call quad(den, "fxn1", lim2) eps=eps scale=1
             cycles= 8 msg="no";
        obj = (theta#num)/den;
return(obj);
finish;

start fxn3(theta) global(alpha, lambda, rho, subdata, i,
                 j, b0, bi, xres);
        num = fxn1(theta);
        eps = 1E-10;
        lim2 = {0 .P};
        call quad(den, "fxn1", lim2) eps=eps scale=1
             cycles= 8 msg="no";
        obj = ((theta-xres)##2)#num/den;
return(obj);
finish;

/*Intials*/
N = 232;
use u.biwei;
read all into bi;
bi = bi[, 2];
ebtheta = repeat(., M);
ebstd = repeat(., M);
b0 = {-3.8169 -0.1090 21.6410 0.249 0.9356 0.8787};
p = 0;
alpha = b0[3];
d = b0[4]; /*variance of random effect*/
lambda = b0[5];
rho = b0[6];
do i = 1 to N;
        index = t(loc(data[, 1] = i));
        subdata = data[index, ];
        ni = nrow(subdata);
                do j = 1 to ni;
```

```
                                        p = p + 1;
                                        eps = 1E-10;
                                        lim = {0 .P};
                                        call quad(xres, "fxn2", lim) eps=
                                             eps scale=1 cycles= 8 msg=
                                             "no";
                                        call quad(std, "fxn3", lim) eps=
                                             eps scale=1 cycles= 8 msg=
                                             "no";
                                        ebtheta[p] = xres;
                                        ebstd[p] = sqrt(std);
                                        *end;
                                end;
        end;
        out = ebtheta||ebstd;
        create u.mythetacmwei from out[colname={'est' 'std'}];
        append from out;
quit;

/*Prediction*/
data allepi;
        merge asthma.asthma1 u.biwei;
        by Patid;
run;
data allepi;
        merge allepi u.mythetacmwei;
run;

data u.AllResultWei;
        set allepi;
        predkappa = est*exp(-4.1993 - 0.0887*(Drug=1) + estimate);
        keep Patid Drug Time estimate StdErrPred est std
             predkappa ;
run;
```

## B.4 The Alternating Imputation Posterior Algorithm

Clayton and Rasbash (1999) suggested a special kind of MCMC algorithm for GLMs with crossed random effects. This algorithm is based on the imputation posterior (IP) algorithm of Tanner and Wong (1987), and will be treated in the context of random-effects models.

Let $\mathbf{Y}$ denote the observed data, $\mathbf{b}$ present the random effects and the fixed model parameters are given by $\boldsymbol{\xi}$. The IP algorithm, which iterates between an 'imputation'(I) step and a 'posterior'(P) step, is similar to Gibbs sampling except that, in the P-step, the whole parameter vector is sampled from its conditional posterior distribution given the

random effects, instead of single components. At iterative step $(t)$, the algorithm can be outlined as follows:

1. I-step: Draw a sample $\mathbf{b}^{(t)}$ from the posterior distribution of $\mathbf{b}$, given $\mathbf{Y}$ and $\boldsymbol{\xi}^{(t-1)}$;
2. P-step: Draw a sample $\boldsymbol{\xi}^{(t)}$ from the posterior distribution of $\boldsymbol{\xi}$, given $\mathbf{Y}$ and $\mathbf{b}^{(t)}$.

As in Gibbs sampling, the algorithm is run until the stationary distribution has been reached (for a burn-in period), say $T$. The overall parameter $(\boldsymbol{\xi}, \mathbf{b})$ is estimated by the mean

$$(\mathrm{E}(\boldsymbol{\xi} \mid \mathbf{Y}); \mathrm{E}(\mathbf{b} \mid \mathbf{Y})) = \left( \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\xi}^{(t)}; \frac{1}{T} \sum_{t=1}^{T} \mathbf{b}^{(t)} \right). \tag{B.26}$$

A few comments are in place. First, in the I-step, the usual EB posterior distribution of the random effects for fixed parameters is used (Skrondal and Rabe-Hesketh, 2009), but with parameters set equal to $\boldsymbol{\xi}^{(t)}$ instead of the maximum likelihood estimates. Second, in the P-step, the random effects drawn from the previous iteration step $(t-1)$ are treated as fixed offsets. The posterior distribution of the parameters is then approximated by a multivariate normal distribution with mean equal to the MLE $\widehat{\boldsymbol{\xi}}^{(t)}$ (treated $\mathbf{b}^{(t)}$ as offsets) and covariance matrix $\widehat{\Sigma}^{(t)}$ derived from the Hessian. This approximate 'sampling distribution'approximates the true Bayesian posterior if uniform priors are assumed for all parameters $\boldsymbol{\xi}$. Using Rao-Blackwellization (Gelfand and Smith, 1990), the variance of the parameter estimates $\boldsymbol{\xi}$ is then estimated by

$$\frac{1}{T} \sum_{t=1}^{T} \widehat{\Sigma}^{(t)} + \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\xi} - \mathrm{E}(\boldsymbol{\xi} \mid \mathbf{Y})) \cdot (\boldsymbol{\xi} - \mathrm{E}(\boldsymbol{\xi} \mid \mathbf{Y}))', \tag{B.27}$$

Clayton and Rasbash (1999) argue that ideally, the P-step should consist of two parts: (1) use restricted maximum likelihood (Bartlett, 1937) to obtain an approximate posterior for the variance and covariance parameters and draw sample from this posterior and (2) approximate the posterior of the fixed parameters by a multivariate normal distribution with mean and variance from the ML solution setting the variance parameters equal to the draws from (1).

Unlike MCMC, the AIP algorithm does not require specification of prior distributions for the model parameters. Furthermore, the algorithm typically converges much more rapidly because the model parameters are updated simultaneously. As pointed out by Clayton and Rasbash (1999), this algorithm also requires many fewer draws than a Gibbs sampler based on scalar nodes to estimate posteriors accurately. The reason for this is that characteristics of the joint posterior distribution can be estimated using Rao-Blackwellization.

# Chapter C

# Supplementary Material for Chapter 5

## C.1 SAS code for implementing the MMM of the WGN model

```
/*************************************************************
SOFTWARE: SAS 9.3.
OBJECTIVE: Implementation of the MMM for WGN model (B.22)-(B.25)
for analyzing the Asthma dataset with full likelihood with
incorporating censoring;
DATASET: Example 9 of Duchateau & Janssen (2008);
VARIABLE DESCRIPTION:
Patid: Patient ID;
Begin and End: time interval between events for each patient;
Status: Right censoring indicator (1 = Asthma Attack, 0 = Censo-
red);
Drug: Treatment indicator (1 = Drug, 0 = Placebo).
REFERENCE: Efendi, Molenberghs and Iddi (2013).
*************************************************************/
/*Weibull-Gamma-Normal model - full likelihood with right censo-
ring - MMM*/
proc nlmixed data = asthma.asthma1 tech = quanew qpoints = 50
             maxit = 1000;
          bounds lambda > 0, alpha > 0;
          parms Beta_0 = -2 Beta_1 = -0.16 lambda = 1 rho = 1
                alpha = 3.3 sigma = 1;
          eta = (Beta_0 + b) + Beta_1*(Drug=1) - sigma/2;
          expeta = exp(eta);
          c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
          c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha) + (rho
               -1)*log(Time) + eta - (alpha+1)*log(lambda*(Time**
               rho)*expeta + alpha);
          loglik  = (status=0)*log(c0) + (status=1)*c1;
```

```
        model Time ~ general(loglik);
        random b ~ normal(0, sigma**2) subject=Patid;
        estimate 'Variance of R.E.s' sigma**2;
run;
```

# Chapter D

# Supplementary Material for Chapter 6

## D.1 Derivation of the Joint Marginal Density

### D.1.1 Case 1: Longitudinal Continous and Repeated Time-to-Event Data

Consider the conditional joint distribution given by (6.6). Integrating over the normal random effects, by using the expansion

$$e^{-\lambda \cdot t_{ik}^{\rho} \cdot \theta_{ik} \cdot e^{\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i}} = \sum_{m_k=0}^{+\infty} \frac{(-1)^{m_k}}{m_k!} \cdot \lambda^{m_k} \cdot t_{ik}^{m_j \cdot \rho} \cdot \theta_{ik}^{m_k} \cdot e^{m_k \cdot (\mathbf{x}_{ik}^{'} \cdot \boldsymbol{\xi}_1 + \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i)} \tag{D.1}$$

and rewrite the expression $[\mathbf{y}_i - (X_i \cdot \boldsymbol{\xi}_2 + Z_i \cdot \mathbf{b}_i)]^{'} \cdot \Sigma_i^{-1} \cdot [\mathbf{y}_i - (X_i \cdot \boldsymbol{\xi}_2 + Z_i \cdot \mathbf{b}_i)]$ by

$$(\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2)^{'} \cdot \Sigma_i^{-1} \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2) - 2 \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2)^{'} \cdot \Sigma_i^{-1} \cdot (Z_i \cdot \mathbf{b}_i) + \mathbf{b}_i^{'} \cdot Z_i^{'} \cdot \Sigma_i^{-1} \cdot Z_i \cdot \mathbf{b}_i, \tag{D.2}$$

and $\sum_{k=1}^{p_i} (m_k+1) \cdot \mathbf{w}_{ik}^{'} \cdot \mathbf{b}_i - 2 \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2)^{'} \cdot \Sigma_i^{-1} \cdot (Z_i \cdot \mathbf{b}_i) - \frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot [(D^{-1} + Z_i^{'} \cdot \Sigma_i^{-1} \cdot Z_i)^{-1}]^{-1} \cdot \mathbf{b}_i$ by

$$-\frac{1}{2} \cdot (\mathbf{b}_i - K)^{'} \cdot [(D^{-1} + Z_i^{'} \cdot \Sigma_i^{-1} \cdot Z_i)^{-1}]^{-1} \cdot (\mathbf{b}_i - K) + L, \tag{D.3}$$

with

$$K = (D^{-1} + Z_i^{'} \cdot \Sigma_i^{-1} \cdot Z_i)^{-1} \cdot S^{'}, \qquad L = \frac{1}{2} \cdot S \cdot (D^{-1} + Z_i^{'} \cdot \Sigma_i^{-1} \cdot Z_i)^{-1} \cdot S^{'},$$

$$S = \sum_{k=1}^{p_i} (m_k + 1) \cdot \mathbf{w}_{ik}^{'} - 2 \cdot (\mathbf{y}_i - X_i \cdot \boldsymbol{\xi}_2)^{'} \cdot \Sigma_i^{-1} \cdot Z_i,$$

the joint distribution, conditional on the gamma random effects, takes the form:

$$f(\mathbf{y}_i,\mathbf{t}_i \mid \boldsymbol{\theta}_i)=\frac{1}{(2\cdot\pi)^{n_i/2}\cdot \mid \Sigma_i^{-1}\mid^{1/2}}\cdot e^{-\frac{1}{2}\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)'\cdot\Sigma_i^{-1}\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)}\cdot\sum_{(l_1,\ldots,l_{p_i})}\left\{\frac{(-1)^{\sum_{k=1}^{p_i}l_k}}{\prod_{k=1}^{p_i}l_k!}\right.$$
$$\left.\cdot e^{\sum_{k=1}^{p_i}(l_k+1)\cdot\mathbf{x}_{ik}'\cdot\boldsymbol{\xi}_1}\cdot\left(\prod_{k=1}^{p_i}\theta_{ik}^{l_k+1}\lambda^{l_k+1}\cdot\rho\cdot t_{ik}^{\rho\cdot l_k+\rho-1}\right)\cdot a_{lk}\right\}\cdot\frac{1}{\mid D\mid^{1/2}}$$
$$\cdot\mid(D^{-1}+Z_i'\cdot\Sigma_i^{-1}\cdot Z_i)^{-1}\mid^{1/2}, \tag{D.4}$$

where

$$a_{lk}=e^{\frac{1}{2}\cdot\left[\sum_{k=1}^{p_i}(l_k+1)\cdot\mathbf{w}_{ik}'-2\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)'\cdot\Sigma_i^{-1}\cdot Z_i\right]\cdot\left(D^{-1}+Z_i'\cdot\Sigma_i\cdot Z_i\right)^{-1}\cdot\left[\sum_{k=1}^{p_i}(l_k+1)\cdot\mathbf{w}_{ik}'-2\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)'\cdot\Sigma_i^{-1}\cdot Z_i\right]'}.$$

Knowing that

$$b_{lk}\equiv\int_\theta\prod_{k=1}^{p_i}\frac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot\theta_{ik}^{l_k+\alpha}\cdot e^{\frac{\theta_{ik}}{\beta}}\cdot d\theta_{ik}=\left[\beta^{l_k-1}\cdot(l_k+\alpha)\cdot(l_k+\alpha-1)\cdot\ldots\cdot\alpha\right]^{p_i}, \tag{D.5}$$

the joint distribution, i.e., by integrating out the Gamma random effects in $(D.4)$, follows to be

$$f(\mathbf{y}_i,\mathbf{t}_i)=\frac{1}{(2\cdot\pi)^{n_i/2}\cdot \mid \Sigma_i^{-1}\mid^{1/2}}\cdot e^{-\frac{1}{2}\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)'\cdot\Sigma_i^{-1}\cdot(\mathbf{y}_i-X_i\cdot\boldsymbol{\xi}_2)}\cdot\sum_{(l_1,\ldots,l_{p_i})}\left\{\frac{(-1)^{\sum_{k=1}^{p_i}l_k}}{\prod_{k=1}^{p_i}l_k!}\right.$$
$$\left.\cdot e^{\sum_{k=1}^{p_i}(l_k+1)\cdot\mathbf{x}_{ik}'\cdot\boldsymbol{\xi}_1}\cdot\left(\prod_{k=1}^{p_i}\lambda^{l_k+1}\cdot\rho\cdot t_{ik}^{\rho\cdot l_k+\rho-1}\right)\cdot a_{lk}\cdot b_{lk}\right\}\cdot\frac{1}{\mid D\mid^{1/2}}$$
$$\cdot\mid(D^{-1}+Z_i'\cdot\Sigma_i^{-1}\cdot Z_i)^{-1}\mid^{1/2}. \tag{D.6}$$

### D.1.2 Case 2: Repeated Binary Outcomes and Repeated Time-to-Event Data

Consider the conditional joint distribution given by (6.15). Specifically, assume that the binary outcome repeatedly measures 1 (similar derivations can be made for other specifications). By using expansion $(D.1)$, denoting $K_{ij}$ by

$$\int_{-\infty}^{\mathbf{x}_{ij}'\cdot\boldsymbol{\xi}_2}\frac{1}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}\cdot(s+\mathbf{z}_{ij}'\cdot\mathbf{b}_i)^2}\cdot ds \tag{D.7}$$

and write the expression $\mathbf{b}_i'\cdot D^{-1}\cdot\mathbf{b}_i+(s+\mathbf{z}_{ij}'\cdot\mathbf{b}_i)^2-2\cdot\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}'\cdot\mathbf{b}_i$ by

$$(\mathbf{b}_i-K)'\cdot[(D^{-1}+\mathbf{z}_{ij}\cdot\mathbf{z}_{ij}')^{-1}]^{-1}\cdot(\mathbf{b}_i-K)+L, \tag{D.8}$$

with

$$K=-(D^{-1}+\mathbf{z}_{ij}\cdot\mathbf{z}_{ij}')^{-1}\cdot\left[\mathbf{z}_{ij}\cdot s-\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right],$$

$$L=s^2-\left[\mathbf{z}_{ij}\cdot s-\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right]'\cdot(D^{-1}+\mathbf{z}_{ij}\cdot\mathbf{z}_{ij}')^{-1}\cdot\left[\mathbf{z}_{ij}\cdot s-\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right],$$

the integration of the normal random effects can be rewritten by

$$\frac{1}{\mid D \mid^{1/2}} \cdot \left(\frac{\alpha_1}{\alpha_1+\beta_1}\right)^{n_i} \cdot \sum_{l_k=0}^{+\infty} \frac{(-1)^{\sum_{k=1}^{p_i} l_k}}{\prod_{k=1}^{p_i} l_k!} \cdot \left(\prod_{k=1}^{p_i} \theta_{ik}^{l_k+1}\right) \cdot e^{\sum_{k=1}^{p_i}(l_k+1)\cdot x_{ik}^{'}\cdot \boldsymbol{\xi}_2}$$
$$\cdot \prod_{k=1}^{p_i} \lambda^{l_k+1} \cdot \rho \cdot t_{ik}^{\rho\cdot l_k+\rho-1} \cdot \prod_{j=1}^{n_i} \frac{1}{\mid D^{-1}+\mathbf{z}_{ij}\cdot \mathbf{z}_{ij}^{'} \mid^{\frac{1}{2}}} \cdot \underbrace{\int_{-\infty}^{\mathbf{x}_{ij}^{'}\cdot\boldsymbol{\xi}_2} \frac{1}{\sqrt{2\cdot\pi}} \cdot e^{-\frac{1}{2}\cdot L} \cdot ds}_{(*)}. \tag{D.9}$$

With the use of a 'completion-of-squares' approach (Wrestler, 1989) and by applying the substitution

$$u = (s + \frac{m_2}{2\cdot m_1}) \cdot m_1^{\frac{1}{2}}, \tag{D.10}$$

with

$$m_1 = 1 - \mathbf{z}_{ij}^{'} \cdot (D^{-1} + \mathbf{z}_{ij} \cdot \mathbf{z}_{ij}^{'})^{-1} \cdot \mathbf{z}_{ij},$$
$$m_2 = \mathbf{z}_{ij}^{'} \cdot (D^{-1} + \mathbf{z}_{ij} \cdot \mathbf{z}_{ij}^{'})^{-1} \cdot \sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik} + \left[\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right]^{'} \cdot (D^{-1} + \mathbf{z}_{ij} \cdot \mathbf{z}_{ij}^{'})^{-1} \cdot \mathbf{z}_{ij},$$

expression $(*)$ can easily be expressed by

$$m_1^{-\frac{1}{2}} \cdot e^{\frac{1}{2}\cdot\left(\frac{m_2^2}{4\cdot m_1}+n_1\right)} \cdot \Phi\left(m_1^{\frac{1}{2}} \cdot \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}_2 + \frac{m_2}{2} \cdot m_1^{-\frac{1}{2}}\right),$$

with

$$n_1 = \left[\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right]^{'} \cdot (D^{-1} + \mathbf{z}_{ij} \cdot \mathbf{z}_{ij}^{'})^{-1} \cdot \left[\sum_{k=1}^{p_i}(l_k+1)\cdot w_{ik}\right].$$

Knowing that $(D.5)$ holds, the joint distribution follows to be

$$f(\mathbf{y}_i = \mathbf{1}, \mathbf{t}_i) = m_1^{-\frac{1}{2}} \cdot e^{\frac{1}{2}\cdot\left(\frac{m_2^2}{4\cdot m_1}+n_1\right)} \cdot \frac{1}{\mid D \mid^{1/2}} \cdot \left(\frac{\alpha_1}{\alpha_1+\beta_1}\right)^{n_i} \cdot \sum_{l_k=0}^{+\infty} \frac{(-1)^{\sum_{k=1}^{p_i} l_k}}{\prod_{k=1}^{p_i} l_k!}$$
$$\cdot e^{\sum_{k=1}^{p_i}(l_k+1)\cdot x_{ik}^{'}\cdot\boldsymbol{\xi}_2} \cdot \prod_{j=1}^{n_i} \frac{1}{\mid D^{-1}+\mathbf{z}_{ij}\cdot\mathbf{z}_{ij}^{'} \mid^{\frac{1}{2}}} \cdot \prod_{k=1}^{p_i} \lambda^{l_k+1} \cdot \rho \cdot t_{ik}^{\rho\cdot l_k+\rho-1}$$
$$\cdot \left[\beta^{l_k-1} \cdot (l_k+\alpha) \cdot (l_k+\alpha-1) \cdot \ldots \cdot \alpha\right]^{p_i}. \tag{D.11}$$

Wrestler, F. E. 1989. Hindu Algebra. In Historical Topics for the Mathematics Classroom; Second edition, edited by John K. Baumgart 1989. Reston: NCTM
Learning and Teaching Mathematics, No.2 Page 7 A Geometrical Introduction to the Method of Completing the Square Anesh Maharaj School of Math. and Statistical Sciences, University of KwaZulu-Natal

## D.2 Derivation of the Correlation Between Both Endpoints and Intraclass Correlation

### D.2.1 Case 3: Bivariate, Repeated Time-to-Event Data

Let $Y_{1ij}$ and $Y_{2ij}$ be the $j$th measurement of cluster (subject) $i$ for time-to-event outcome 1 and 2 respectively. The linear part for the two responses are:

$$\eta_{1ij} = \Delta_{1ij} + \mathbf{z}_{1ij}^{'} \cdot \mathbf{b}_{1i},$$
$$\eta_{2ij} = \Delta_{2ij} + \mathbf{z}_{2ij}^{'} \cdot \mathbf{b}_{2i},$$

with

$$\Delta_{1ij} = -\log(\alpha_{1j} \cdot \beta_{1j}) + \mathbf{x}_{1ij}^{'}\boldsymbol{\xi}_1^m - \frac{\mathbf{z}_{1ij}^{'} \cdot D_{11} \cdot \mathbf{z}_{1ij}}{2},$$
$$\Delta_{2ij} = -\log(\alpha_{2j} \cdot \beta_{2j}) + \mathbf{x}_{2ij}^{'}\boldsymbol{\xi}_2^m - \frac{\mathbf{z}_{2ij}^{'} \cdot D_{22} \cdot \mathbf{z}_{2ij}}{2},$$
$$\begin{bmatrix}\mathbf{b}_{1i}\\ \mathbf{b}_{2i}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}D_{11} & D_{12}\\ D_{12} & D_{22}\end{bmatrix}\right)$$

For simplicity, the random-intercept approach is chosen, where censoring is not present in both outcomes. The bivariate normal distribution (6.24) is used for derivations. At last, $\alpha_v \cdot \beta_v = 1$ is assumed for reasons of identifiability ($v = 1, 2$).

1. *Derivation of the correlation between both endpoints*

   The correlation between the two outcomes is, by definition:

   $$\text{Corr}(Y_{1ij}, Y_{2ij}) = \frac{\text{Cov}(Y_{1ij}, Y_{2ij})}{\sqrt{\text{Var}(Y_{1ij})} \cdot \sqrt{\text{Var}(Y_{2ij})}}. \tag{D.12}$$

   Knowing that

   $$\text{Cov}(Y_{1ij}, Y_{2ij}) = \text{E}[\text{Cov}(Y_{1ij}, Y_{2ij} \mid b_{1i}, b_{2i})] + \text{Cov}[\text{E}(Y_{1ij} \mid b_{1i}, b_{2i}), \text{E}(Y_{2ij} \mid b_{1i}, b_{2i})].$$

   and the fact that $\text{Cov}(Y_{1ij}, Y_{2ij} \mid b_{1i}, b_{2i}) = 0$, expression $(D.12)$ can easily be re-written by

   $$\text{Corr}(Y_{1ij}, Y_{2ij}) = \frac{\text{Cov}[\text{E}(Y_{1ij} \mid b_{1i}), \text{E}(Y_{2ij} \mid b_{2i})]}{\sqrt{\text{Var}(Y_{1ij})} \cdot \sqrt{\text{Var}(Y_{2ij})}}. \tag{D.13}$$

   By integrating the gamma random effect, it can easily be shown, for $v = 1, 2$, that

   $$f(y_{vij} \mid b_{vi}) = \frac{\lambda_v \cdot e^{\Delta_{vij}+b_{vi}} \cdot \rho_v \cdot y_{vij}^{\rho_v - 1} \cdot \alpha_v^{\alpha_v+1}}{(\alpha_v + \lambda_v \cdot y_{vij}^{\rho_v} \cdot e^{\Delta_{vij}+b_{vi}})^{\alpha_v+1}}. \tag{D.14}$$

Therefore, the conditional expectation for outcome $v = 1, 2$ is given by

$$\mathrm{E}(Y_{vij} \mid b_{vi}) = \int_0^{+\infty} y_{vij} \cdot f(y_{vij} \mid b_{vi}) \cdot dy_{vij} = \frac{\alpha_v^{1/\rho_v} \cdot B\left(\alpha_v - \frac{1}{\rho_v}, \frac{1}{\rho_v}\right)}{\rho_v \cdot (\lambda_v \cdot e^{\Delta_{vij} + b_{vi}})^{1/\rho_v}}. \tag{D.15}$$

Using formula $(D.15)$, the covariance between $\mathrm{E}(Y_{1ij} \mid b_{1i})$ and $\mathrm{E}(Y_{2ij} \mid b_{2i})$, i.e., the numerator of $(D.13)$, can easily be formulated to

$$\begin{aligned}
\mathrm{Cov}[\mathrm{E}(Y_{1ij} \mid b_{1i}), \mathrm{E}(Y_{2ij} \mid b_{2i})] &= \frac{\alpha_1^{1/\rho_1} \cdot \alpha_2^{1/\rho_2} \cdot B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1}\right) \cdot B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)}{\rho_1 \cdot (\lambda_1 \cdot e^{\Delta_{1ij}})^{1/\rho_1} \cdot \rho_2 \cdot (\lambda_2 \cdot e^{\Delta_{2ij}})^{1/\rho_2}} \\
&\quad \cdot \mathrm{Cov}\left(e^{-b_{1i}/\rho_1}, e^{-b_{2i}/\rho_2}\right) \\
&= \frac{\alpha_1^{1/\rho_1} \cdot \alpha_2^{1/\rho_2} \cdot B\left(\alpha_1 - \frac{1}{\rho_1}, \frac{1}{\rho_1}\right) \cdot B\left(\alpha_2 - \frac{1}{\rho_2}, \frac{1}{\rho_2}\right)}{\rho_1 \cdot (\lambda_1 \cdot e^{\Delta_{1ij}})^{1/\rho_1} \cdot \rho_2 \cdot (\lambda_2 \cdot e^{\Delta_{2ij}})^{1/\rho_2}} \\
&\quad \cdot \left(e^{\frac{r \cdot d_1 \cdot d_2}{\rho_1 \cdot \rho_2}} - 1\right) \cdot e^{\frac{1}{2} \cdot \left(\frac{d_1^2}{\rho_1^2} + \frac{d_2^2}{\rho_2^2}\right)}.
\end{aligned} \tag{D.16}$$

In addition, the variance of $Y_{vij}$, for $v = 1, 2$, equals to

$$\mathrm{Var}(Y_{vij}) = \frac{\alpha_v^{1/\rho_v} \cdot e^{d_v^2/\rho_v^2} \cdot \left[B\left(\alpha_v - \frac{2}{\rho_v}, \frac{2}{\rho_v}\right) \cdot e^{d_v^2/\rho_v^2} - B\left(\alpha_v - \frac{2}{\rho_v}, \frac{2}{\rho_v}\right)^2\right]}{\rho_v \cdot (\lambda_v \cdot e^{\Delta_{vij}})^{1/\rho_v}}. \tag{D.17}$$

Imputing $(D.16)$ and $(D.17)$ in $(D.13)$ gives formula (6.26) for the correlation between the two outcomes.

2. *Derivation of the intraclass correlation*

By definition, the correlation between the $j$th and $k$th measurements of cluster (subject) $i$ for outcome $v$ ($v = 1, 2$) is

$$\mathrm{Corr}(Y_{vij}, Y_{vik}) = \frac{\mathrm{Cov}(Y_{vij}, Y_{vik})}{\sqrt{\mathrm{Var}(Y_{vij})} \cdot \sqrt{\mathrm{Var}(Y_{vik})}}. \tag{D.18}$$

Given that the random effect between the measurements are independent, i.e., where $\mathrm{Cov}(Y_{vij}, Y_{vik} \mid b_{vi}) = 0$ holds, for $v = 1, 2$, formula $(D.18)$ can simply be rewritten by

$$\mathrm{Corr}(Y_{vij}, Y_{vik}) = \frac{\mathrm{Cov}[\mathrm{E}(Y_{vij} \mid b_{vi}), \mathrm{E}(Y_{vik} \mid b_{vi})]}{\sqrt{\mathrm{Var}(Y_{vij})} \cdot \sqrt{\mathrm{Var}(Y_{vik})}}. \tag{D.19}$$

From formula $(D.15)$, the numerator of $(D.19)$ can be expressed by

$$\mathrm{Cov}[\mathrm{E}(Y_{vij} \mid b_{vi}), \mathrm{E}(Y_{vik} \mid b_{vi})] = \frac{\alpha_v^{2/\rho_v} \cdot B\left(\alpha_v - \frac{1}{\rho_v}, \frac{1}{\rho_v}\right)^2}{\rho_v^2 \cdot (\lambda_v^2 \cdot e^{\Delta_{vij}+\Delta_{vik}})^{1/\rho_v}} \cdot \left(e^{d_v^2/\rho_v^2} - 1\right) \cdot e^{d_v^2/\rho_v^2}. \tag{D.20}$$

By imputing expressions $(D.17)$ and $(D.20)$ in $(D.19)$, formula (6.25) is achieved for the intraclass correlation.

# Chapter E

# Supplementary Material for Chapter 8

## E.1 Derivation to achieve the proportional relationship between the directional derivative of the log-likelihood $\ell(\cdot)$ at $G$ into the direction $H$ and the gradient function $\Delta(G, \mathbf{b})$

In order to achieve the proportional relationship between the directional derivative of the log-likelihood $\ell(\cdot)$ at $G$ into the direction $H$ and the gradient function $\Delta(G, \mathbf{b})$, the following explicit derivations have been made by Verbeke and Molenberghs (2013):

$$\begin{aligned}
\Phi(G, H) &= \left. \frac{\partial \sum_{i=1}^{N} \ln[(1-\alpha) \cdot f_i(\mathbf{y}_i \mid G) + \alpha \cdot f_i(\mathbf{y}_i \mid H)]}{\partial \alpha} \right|_{\alpha=0} \\
&= \sum_{i=1}^{N} \frac{f_i(\mathbf{y}_i \mid H) - f_i(\mathbf{y}_i \mid G)}{f_i(\mathbf{y}_i \mid G)} \\
&= \sum_{i=1}^{N} \frac{f_i(\mathbf{y}_i \mid H)}{f_i(\mathbf{y}_i \mid G)} - N \\
&= \sum_{i=1}^{N} \frac{\int_{\mathbf{b}} f_i(\mathbf{y}_i \mid \mathbf{b}) \cdot dH(\mathbf{b})}{f_i(\mathbf{y}_i \mid G)} - N \\
&= N \cdot \int_{\mathbf{b}} \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{f_i(\mathbf{y}_i \mid \mathbf{b})}{f_i(\mathbf{y}_i \mid G)} \cdot dH(\mathbf{b}) - N \\
&= N \cdot \left[ \int_{\mathbf{b}} \Delta(G, \mathbf{b}) \cdot dH(\mathbf{b}) - 1 \right] \\
&\propto \int_{\mathbf{b}} \Delta(G, \mathbf{b}) \cdot dH(\mathbf{b}) - 1. \qquad \text{(E.1)}
\end{aligned}$$

# Chapter F

# Supplementary Material for Chapter 9

Lesaffre and Verbeke (1998) introduced an influence assessment paradigm for the linear mixed model, by following the approach of local influence proposed by Cook (1986). Their developments will be presented here (Section F.1.1). Additionally, Rakhmawati et al (2014) derived an alternative derivation on the likelihood in integral form (Section F.1.2). This paradigm is used as starting point in the developments of Rakhmawati et al (2014) influence assessment paradigm for the combined model of Section 3.4. Special attention is given on the local influence for the Weibull-normal model (Section F.2). Derivations for other models were made as well, and can be found in Rakhmawati et al (2014).
Cook 1986, Journal of the Royal Statistical Society, Series B 48, 133-169)

## F.1 Local Influence for the Linear Mixed Model

### F.1.1 Standard approach, based on the Marginal Likelihood

To stay in line with Lesaffre and Verbeke (1998), the marginal linear mixed model (3.24) is considered, with conditional independence assumption $\Sigma_i = \sigma^2 \cdot I_{n_i}$. Here, $I_{n_i}$ denotes the $n_i$ x $n_i$ identity matrix.

For $C_i$ as in formula (9.7), a convenient form can be derived:

$$C_i = -2 \cdot \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^1_{(i)}\right)' \cdot \ddot{L}_{(i)} \cdot \ddot{L}^{-1} \cdot \ddot{L}_{(i)} \cdot \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}^1_{(i)}\right), \tag{F.1}$$

where a subscript $(i)$ indicates that the corresponding quantity is based on the deletion of the $i$th subject and further the vector $\widehat{\boldsymbol{\theta}}^1_{(i)}$ is the one-step approximation to $\widehat{\boldsymbol{\theta}}_{(i)}$ obtained from a single Newton-Raphson step in the maximization procedure of $l_{(i)}(\boldsymbol{\theta})$, using $\widehat{\boldsymbol{\theta}}$ as the starting value. For sufficiently large sample size, it follows that $C_i$ is an approximation to the classical global case-deletion diagnostics. Note that the expression is exact when properly used for local influence purposes.

It is advantageous that $C_i$ admits a closed form (9.7). Lesaffre and Verbeke (1998) decomposed $C_i$ into five interpretable components. Let $\boldsymbol{\mathcal{R}}_i$, $\mathcal{X}_i$, and $\mathcal{Z}_i$ denote the "standardized" residuals and covariates for the $i$th individual, defined by $\boldsymbol{\mathcal{R}}_i = V_i^{-1/2} \cdot \mathbf{r}_i$,

$\mathcal{X}_i = V_i^{-1/2} \cdot X_i$, and $\mathcal{Z}_i = V_i^{-1/2} \cdot Z_i$, respectively, with $\mathbf{r}_i = \mathbf{y}_i - X_i \cdot \widehat{\boldsymbol{\beta}}$. Further, for a matrix $A$, let $\|A\| = \sqrt{\text{tr}(A' \cdot A)}$ be the Frobenius norm of $A$ (Golub and Van Loan, 1989). The interpretable components in $C_i$ are then

$$\|\mathcal{X}_i \cdot \mathcal{X}_i^{'}\|, \qquad \|\boldsymbol{\mathcal{R}}_i\|, \qquad \|\mathcal{Z}_i \cdot \mathcal{Z}_i^{'}\|, \qquad \|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|, \qquad \|V_i^{-1}\|. \tag{F.2}$$

First, $\|\mathcal{X}_i \cdot \mathcal{X}_i^{'}\|$ measures the length of the standardized covariates in the mean structure and $\|\boldsymbol{\mathcal{R}}_i\|$ is an overall measure for how well the observed data for the $i$th subject are predicted by the mean structure $X_i \cdot \widehat{\boldsymbol{\beta}}$. Second, the components $\|\mathcal{Z}_i \cdot \mathcal{Z}_i^{'}\|$ and $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|$ have a similar meaning, but then for the covariance structure. For example, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|$ will be zero only if $V_i$ equals $\mathbf{r}_i \cdot \mathbf{r}_i^{'}$. Note that $\mathbf{r}_i \cdot \mathbf{r}_i^{'}$ is an estimate for $\text{Var}(\mathbf{y}_i)$, which only assumes the mean to be correctly modeled as $X_i \cdot \boldsymbol{\beta}$. Therefore, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|$ can be interpreted as a residual, capturing how well the covariance structure of the data is modeled by $V_i = Z_i \cdot D \cdot Z_i^{'} + \sigma^2 \cdot I_{n_i}$. Finally, the fifth component $\|V_i^{-1}\|$ will be large if $V_i$ has small eigenvalues, which indicates that the $i$th subject is assumed to have small variability.

The decomposition of $C_i$ immediately suggests a practical procedure to find an explanation for the influential nature of an individual, i.e., when $C_i$ is large, we examine the diagnostics. Such plots are useful to graphically inspect the individuals in view of their influential nature. Thus, it is sensible to start with an index plot of $C_i$. Following this, the index plots of $(F.2)$ can be examined. A recurrent practical difficulty with diagnostics is to establish a threshold above which an individual is defined as remarkable. It follows from (9.7) that

$$\sum_{i=1}^{N} C_i = -2 \cdot \text{tr}\left(\ddot{L}^{-1} \cdot \sum_{i=1}^{N} \Delta_i \cdot \Delta_i^{'}\right), \tag{F.3}$$

which converges to $2 \cdot s$, for $N$ approaching infinity. As for leverage in linear regression (Neter, Wasserman and Kutner 1990, pp. 395-396), one could classify an individual for which $C_i$ is larger than twice the average value (larger than $4 \cdot s/N$, for $N$ large) as being influential. However, unlike for the leverage situation, $2 \cdot s$ is only the approximate sum of the $C_i$, which will not be accurate if the model is not correctly specified (such that $\ddot{L}^{-1} \cdot \sum_{i=1}^{N} \Delta_i \cdot \Delta_i^{'}$ does not converge to $I_s$) or if $N$ is too small for the asymptotic results to yield good approximations. In such cases, Lesaffre and Verbeke (1998) proposed to replace $2 \cdot s$ by the actual sum, and we call the $i$th subject influential if $C_i$ is larger than the cutoff value $2 \cdot \sum_{i=1}^{N} C_i/N$.

Given decomposition result $(F.2)$, it is interesting to consider sub-vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of fixed effects and variance components, respectively, with corresponding influences $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{\alpha})$, respectively. Given that the fixed effects and variance components are asymptotically independent, it follows that

$$C_i \approx C_i(\boldsymbol{\beta}) + C_i(\boldsymbol{\alpha}). \tag{F.4}$$

Lesaffre and Verbeke (1998) further showed that $C_i(\boldsymbol{\beta})$ can be decomposed using only the first two components $\|\mathcal{X}_i \cdot \mathcal{X}_i^{'}\|$ and $\|\boldsymbol{\mathcal{R}}_i\|$, while the last three components $\|\mathcal{Z}_i \cdot \mathcal{Z}_i^{'}\|$, $\|I - \boldsymbol{\mathcal{R}}_i \cdot \boldsymbol{\mathcal{R}}_i^{'}\|$ and $\|V_i^{-1}\|$ feature in the decomposition of $C_i(\boldsymbol{\alpha})$. Asymptotically therefore, influence for the fixed effects and for the variance components can be scrutinized by studying the first two and the last three interpretable components, respectively.

### F.1.2 Integral-based expression

As previewed in Section 4, the integral-based approach is used here as an alternative way to alleviate complexities with the explicit marginal likelihood expressions. To prepare for developments of Poisson, probit, logit and Weibull cases, we set out this way for the linear mixed model.

The marginal density corresponding to the linear mixed model is defined by the following expression:

$$\tilde{f}(\mathbf{y}_i) = \int \tilde{f}(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i) \cdot \tilde{f}(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i. \tag{F.5}$$

The conditional density of the response variable takes the form:

$$\begin{aligned}\tilde{f}(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i) &= \left(\frac{1}{2 \cdot \pi \cdot \sigma^2}\right)^{n_i/2} \cdot \exp\left[-\frac{1}{2 \cdot \sigma^2} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot \mathbf{b}_i)' \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot \mathbf{b}_i)\right] \\ &= (2 \cdot \pi \cdot s)^{-n_i/2} \cdot \exp\left[f(\mathbf{y}_i)\right], \end{aligned} \tag{F.6}$$

where $f(\mathbf{y}_i) = -(2 \cdot s)^{-1} \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i)' \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i)$, $\widehat{\mathbf{y}}_i = \mathbf{X}_i \cdot \boldsymbol{\beta} + \mathbf{Z}_i \cdot \mathbf{b}_i$ and $s = \sigma^2$. The conditional density of the normal random effect is:

$$f(\mathbf{b}_i) = \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot \mathbf{b}_i' \cdot D^{-1} \cdot \mathbf{b}_i} = (2 \cdot \pi)^{-q/2} \cdot \mid D \mid^{-1/2} \cdot \exp\left[g(\mathbf{b}_i)\right], \tag{F.7}$$

where $g(\mathbf{b}_i) = -\frac{1}{2} \cdot \mathbf{b}_i' \cdot D^{-1} \cdot \mathbf{b}_i$. Thus, the marginal density for the linear mixed model is:

$$\tilde{f}(\mathbf{y}_i) = (2 \cdot \pi)^{-(n_i+q)/2} \cdot s^{-n_i/2} \cdot \mid D \mid^{-1/2} \cdot \int \exp\left[f(\mathbf{y}_i) + g(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i. \tag{F.8}$$

From $(F.8)$ the likelihood derives as:

$$L(\boldsymbol{\beta}, D, s) = \prod_{i=1}^{N} \tilde{f}(\mathbf{y}_i), \tag{F.9}$$

and the corresponding log-likelihood is (9.1). Thus, the log-likelihood contribution of the $i$th individual takes the form:

$$\begin{aligned} l_i(\boldsymbol{\beta}, D, s) &= \log\{(2 \cdot \pi)^{-(n_i+q)/2} \cdot s^{-n_i/2} \cdot \mid D \mid^{-1/2} \cdot \int \exp\left[f(\mathbf{y}_i) + g(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i\} \\ &= -\frac{n_i + q}{2} \cdot \log(2 \cdot \pi) - \frac{n_i}{2} \cdot \log(s) - \frac{1}{2} \cdot \log \mid D \mid + \log\{\int \exp\left[f(\mathbf{y}_i) + g(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i\} \\ &\propto -\frac{n_i}{2} \cdot \log(s) - \frac{1}{2} \cdot \log \mid D \mid + \log K_i, \end{aligned} \tag{F.10}$$

where $K_i = \int I_i \cdot d\mathbf{b}_i$ and $I_i = \exp\left[f(\mathbf{y}_i) + g(\mathbf{b}_i)\right]$.

To derive the local influence as described in (9.7), the components of local influence need to be derived. Lesaffre and Verbeke (1998) showed that $C_i$ equals:

$$C_i = 2 \cdot \|\ddot{L}^{-1}\| \cdot \|\Delta_i\|^2 \cdot \cos(\varphi_i), \tag{F.11}$$

where $\varphi_i$ is the angle between $\text{vec}(-\ddot{L}^{-1})$ and $\text{vec}(\Delta_i \cdot \Delta_i^{'})$, $\Delta_i$ is the first derivative of $l_i(\boldsymbol{\beta}, D, s)$ with respect to the model parameters, and $\ddot{L}^{-1}$ is the $s?s$ matrix of second derivatives of $l(\boldsymbol{\beta}, D, s)$ with respect to the parameters.

The procedure to construct derivatives with respect to the parameters is as follows. First, the derivative with respect to fixed effect $\boldsymbol{\beta}$ is:

$$\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial \boldsymbol{\beta}} = \frac{1}{K_i} \cdot \int I_i \cdot \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i) \cdot d\boldsymbol{\beta} = \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot \frac{\mathbf{L}_i}{K_i}, \tag{F.12}$$

where

$$K_i = \int I_i \cdot d\mathbf{b}_i = \int \exp\left[f(\mathbf{y}_i) + g(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i = c \cdot \tilde{\phi}(\mathbf{y}_i) \tag{F.13}$$

and

$$\begin{aligned}
\mathbf{L}_i &= \int I_i \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i) \cdot d\mathbf{b}_i \\
&= \int I_i \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot \mathbf{b}_i) \cdot d\mathbf{b}_i \\
&= \mathbf{y}_i \cdot \int I_i \cdot d\mathbf{b}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} \cdot \int I_i \cdot d\mathbf{b}_i + \mathbf{Z}_i \cdot \int I_i \cdot \mathbf{b}_i \cdot d\mathbf{b}_i
\end{aligned} \tag{F.14}$$

Component $\int I_i \cdot \mathbf{b}_i \cdot d\mathbf{b}_i$ of $\mathbf{L}_i$ can be rewritten as:

$$\begin{aligned}
\int I_i \cdot \mathbf{b}_i \cdot d\mathbf{b}_i &= \int c \cdot \tilde{\phi}(\mathbf{y}_i, \mathbf{b}_i) \cdot \mathbf{b}_i \cdot d\mathbf{b}_i \\
&= c \cdot \int \tilde{\phi}(\mathbf{y}_i) \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot \mathbf{b}_i \cdot d\mathbf{b}_i \\
&= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \int \mathbf{b}_i \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot d\mathbf{b}_i \\
&= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot E(\mathbf{b}_i \mid \mathbf{y}_i) \\
&= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{'} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta}) \\
&= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{'} \cdot \mathbf{r}_i,
\end{aligned} \tag{F.15}$$

where $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta}$. Expanding the component functions of $(F.12)$ leads to:

$$\begin{aligned}
\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial \boldsymbol{\beta}} &= \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot \frac{\mathbf{L}_i}{K_i} \\
&= \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot \left[\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta})\right] \\
&= \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot \left[(I_{n_i} - \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{-1}) \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta})\right] \\
&= \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot \left\{\left[(s + \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}) \cdot \mathbf{V}_i^{-1} - \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{-1}\right] \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta})\right\} \\
&= \frac{1}{s} \cdot \mathbf{X}_i^{'} \cdot s \cdot \mathbf{V}_i^{-1} \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta}) \\
&= \mathbf{X}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i.
\end{aligned} \tag{F.16}$$

Second, the derivative with respect to $s \equiv \sigma^2$ is as follows:

$$\begin{aligned}\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial \boldsymbol{s}} &= -\frac{n_i}{s} + \frac{1}{K_i} \cdot \int I_i \cdot \frac{1}{2 \cdot s^2} \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i)' \cdot (\mathbf{y}_i - \widehat{\mathbf{y}}_i) \cdot d\mathbf{b}_i \\ &= -\frac{n_i}{s} - \frac{1}{s \cdot K_i} \cdot \int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i \\ &= -\frac{1}{s} \cdot \left[ \frac{n_i}{2} + \frac{1}{K_i} \cdot \int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i \right], \end{aligned} \tag{F.17}$$

where $K_i$ is given in $(F.13)$. The component $\int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i$ can be rewritten as:

$$\begin{aligned}\int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i &= -\frac{1}{2 \cdot s} \cdot \int I_i \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot \mathbf{b}_i)' \cdot (\mathbf{y}_i - \mathbf{X}_i \cdot \boldsymbol{\beta} - \mathbf{Z}_i \cdot \mathbf{b}_i) \cdot d\mathbf{b}_i \\ &= -\frac{1}{2 \cdot s} \cdot \int I_i \cdot (\mathbf{r}_i - \mathbf{Z}_i \cdot \mathbf{b}_i)' \cdot (\mathbf{r}_i - \mathbf{Z}_i \cdot \mathbf{b}_i) \cdot d\mathbf{b}_i \\ &= -\frac{1}{2 \cdot s} \cdot \left( \mathbf{r}_i' \cdot \mathbf{r}_i \cdot \int I_i \cdot d\mathbf{b}_i - \mathbf{r}_i' \cdot \mathbf{Z}_i \cdot \int \mathbf{b}_i \cdot I_i \cdot d\mathbf{b}_i \right) \\ &\quad - \frac{1}{2 \cdot s} \cdot \left[ - \left( \int \mathbf{b}_i \cdot I_i \cdot d\mathbf{b}_i \right)' \cdot \mathbf{Z}_i' \cdot \mathbf{r}_i + \int \mathbf{b}_i' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot \mathbf{b}_i \cdot I_i \cdot d\mathbf{b}_i \right] \\ &= -\frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left[ \mathbf{r}_i' \cdot \mathbf{r}_i - \mathbf{r}_i' \cdot \mathbf{Z}_i \cdot \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) \right] \\ &\quad - \frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left\{ - \left[ \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) \right]' \cdot \mathbf{Z}_i' \cdot \mathbf{r}_i + \mathrm{E}(\mathbf{b}_i' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot \mathbf{b}_i) \right\}, \end{aligned} \tag{F.18}$$

where $\mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) = D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i$ and

$$\begin{aligned}\mathrm{E}(\mathbf{b}_i' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot \mathbf{b}_i) &= \mathrm{tr} \left[ \mathbf{Z}_i \cdot \mathrm{Var}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot \mathbf{Z}_i' \right] + \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i)' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) \\ &= \mathrm{tr} \left[ \mathbf{Z}_i \cdot (\mathbf{Z}_i' \cdot s^{-1} \cdot \mathbf{Z}_i + D^{-1})^{-1} \right] \\ &\quad + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i \\ &= \mathrm{tr} \left\{ \mathbf{Z}_i \cdot \left[ D - D \cdot \mathbf{Z}_i' \cdot (s + \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i')^{-1} \cdot \mathbf{Z}_i \cdot D \right] \cdot \mathbf{Z}_i' \right\} \\ &\quad + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i \\ &= \mathrm{tr} \left[ \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' + (I_{n_i} - \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i') \right] \\ &\quad + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i \\ &= \mathrm{tr}(\mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot s) + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i \\ &= \mathrm{tr} \left[ (\mathbf{V}_i - s) \cdot \mathbf{V}_i^{-1} \cdot s \right] + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i \\ &= \left[ n_i \cdot s - s^2 \cdot \mathrm{tr}(\mathbf{V}_i^{-1}) \right] + \mathbf{r}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D' \cdot \mathbf{Z}_i' \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i' \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i. \end{aligned} \tag{F.19}$$

Thus, ($F$.18) simplifies to:

$$\begin{aligned}\int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i &= -\frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left[\mathbf{r}_i^{'} \cdot (I_{n_i} - \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}) \cdot \mathbf{r}_i\right] \\ &\quad - \frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left[\mathbf{r}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D^{'} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i + n_i \cdot s \right. \\ &\quad \left. -s^2 \cdot \mathrm{tr}(\mathbf{V}_i^{-1})\right] \\ &= -\frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left\{\mathbf{r}_i^{'} \cdot \left[(I_{n_i} - \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}) \right.\right. \\ &\quad \left.\left. \cdot (I_{n_i} - \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'})^{'}\right] \cdot \mathbf{r}_i\right\} - \frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left[n_i \cdot s - s^2 \cdot \mathrm{tr}(\mathbf{V}_i^{-1})\right] \\ &= -\frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left\{\mathbf{r}_i^{'} \cdot M_i \cdot M_i^{'} \cdot \mathbf{r}_i + \left[n_i \cdot s - s^2 \cdot \mathrm{tr}(\mathbf{V}_i^{-1})\right]\right\} \\ &= -\frac{1}{2 \cdot s} \cdot c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \left[\mathbf{r}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot s \cdot s \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i + n_i \cdot s - s^2 \cdot \mathrm{tr}(\mathbf{V}_i^{-1})\right], \end{aligned} \tag{F.20}$$

where $M_i = \mathbf{V}_i^{-1} \cdot (s + \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}) - \mathbf{V}_i^{-1} \cdot \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'}$.

Expanding the components of ($F$.17) leads to:

$$\begin{aligned}\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial s} &= -\frac{1}{s} \cdot \left[\frac{n_i}{2} + \frac{1}{K_i} \cdot \int I_i \cdot f(\mathbf{y}_i) \cdot d\mathbf{b}_i\right] \\ &= -\frac{n_i}{2 \cdot s} + \frac{1}{2} \cdot \mathbf{r}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i + \frac{n_i}{2 \cdot s} - \frac{1}{2} \cdot \mathrm{tr}(\mathbf{V}_i^{-1}) \\ &= -\frac{1}{2} \cdot \left[\mathrm{tr}(\mathbf{V}_i^{-1}) - \mathbf{r}_i^{'} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{V}_i^{-1} \cdot \mathbf{r}_i\right]. \end{aligned} \tag{F.21}$$

Third, the derivative with respect to $D$ is:

$$\begin{aligned}\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial d_{jk}} &= -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot (D^{-1})_{jk} + \frac{1}{K_i} \cdot \int I_i \cdot \frac{\partial g(\mathbf{b}_i)}{\partial d_{jk}} \cdot d\mathbf{b}_i \\ &= -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot (D^{-1})_{jk} + \frac{1}{2 \cdot K_i} \cdot \int I_i \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \cdot d\mathbf{b}_i. \end{aligned} \tag{F.22}$$

where $d_{jk}$ is the $(j,k)$ element of $D$, and $E_{ij}$ is a matrix of zeros everywhere except a one in entries $(j,k)$ and $(k,j)$. The integral part of the first derivative with respect to $D$ can be written as:

$$\begin{aligned}\int I_i \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \cdot d\mathbf{b}_i &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \int \mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot d\mathbf{b} \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \mathrm{E}(\mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \mid \mathbf{y}_i), \end{aligned} \tag{F.23}$$

where

$$
\begin{aligned}
\mathrm{E}(\mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \mid \mathbf{y}_i) &= \mathrm{E}\left[\mathrm{tr}(\mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i)\right] \\
&= \mathrm{tr}\left\{\left[D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i \mid \mathbf{y}_i)\right]\right\} \\
&= \mathrm{tr}\left\{D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \left[\mathrm{Var}(\mathbf{b}_i \mid \mathbf{y}_i) + \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i)^{'} \cdot \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i)\right]\right\} \\
&= \mathrm{tr}\left[D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathrm{Var}(\mathbf{b}_i \mid \mathbf{y}_i)\right] \\
&\quad + \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i)^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i) \\
&= \mathrm{tr}\left\{D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \left[D - D \cdot \mathbf{Z}_i^{'} \cdot (s + \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'})^{-1} \cdot \mathbf{Z}_i \cdot D\right]\right\} \\
&\quad + \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)} \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i \\
&= \mathrm{tr}\left[D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot (D - D \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i \cdot D)\right] \\
&\quad + (2 - \delta_{jk}) \cdot \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)} \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i \\
&= \mathrm{tr}(D^{-1} \cdot E_{jk}) - \mathrm{tr}(E_{jk} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i \cdot D) \\
&\quad + (2 - \delta_{jk}) \cdot \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)} \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i \\
&= (2 - \delta_{jk}) \cdot (D^{-1})_{jk} - (2 - \delta_{jk}) \cdot \mathbf{Z}_i^{(j)^{'}} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(k)} \\
&\quad + (2 - \delta_{jk}) \cdot \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)} \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i \\
&= (2 - \delta_{jk}) \cdot \left[(D^{-1})_{jk} - \mathbf{Z}_i^{(j)^{'}} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(k)} + \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)}\right. \\
&\quad \left. \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i\right]. \qquad \text{(F.24)}
\end{aligned}
$$

Expanding the components of $(F.22)$ leads to:

$$
\frac{\partial l_i(\boldsymbol{\beta}, D, s)}{\partial d_{jk}} = -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot (\mathbf{Z}_i^{(j)^{'}} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(k)} - \mathbf{r}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{Z}_i^{(j)} \cdot \mathbf{Z}_i^{(k)^{'}} \cdot \mathbf{Z}_i^{'} \cdot \mathbf{V}^{-1} \cdot \mathbf{r}_i). \qquad \text{(F.25)}
$$

This integral-based result, based on $(F.16)$, $(F.21)$, and $(F.25)$ is identical to the standard one of Lesaffre and Verbeke (1998). Hence also, the same interpretable components as in $(F.2)$ ensue.

## F.2 Local Influence for the Weibull-normal Model

The general Weibull model for repeated measurement data as described in (9.14)–(9.15) can be re-expressed as:

$$\begin{aligned} f(\mathbf{y}_i \mid \boldsymbol{\theta}_i, \mathbf{b}_i) &= \prod_{j=1}^{n_i} \lambda \cdot \rho \cdot y_{ij}^{\rho-1} \cdot \exp(\boldsymbol{\mu}_{ij}) \cdot \exp\left[-\lambda \cdot y_{ij}^{\rho} \cdot \exp(\boldsymbol{\mu}_{ij})\right] \\ &= \lambda \cdot \rho \cdot \left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1}\right) \cdot \exp\left\{\sum_{j=1}^{n_i} \left[\boldsymbol{\mu}_{ij} - \lambda \cdot y_{ij}^{\rho} \cdot \exp(\boldsymbol{\mu}_{ij})\right]\right\} \\ &= \lambda \cdot \rho \cdot \left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1}\right) \cdot \exp\left[\tilde{f}(\mathbf{y}_i)\right], \end{aligned} \tag{F.26}$$

where $\boldsymbol{\mu}_{ij} = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i$ and $\tilde{f}(\mathbf{y}_i) = \sum_{j=1}^{n_i} \left[\boldsymbol{\mu}_{ij} - \lambda \cdot y_{ij}^{\rho} \cdot \exp(\boldsymbol{\mu}_{ij})\right]$. Thus, the marginal density of the Weibull-model takes the following form:

$$\begin{aligned} f(\mathbf{y}_i) &= \int f(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i \\ &= \frac{\lambda \cdot \rho \cdot (\prod_{j=1}^{n_i} y_{ij}^{\rho-1})}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot \int \exp\left[\tilde{f}(\mathbf{y}_i) + \tilde{g}(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i, \end{aligned} \tag{F.27}$$

where $\tilde{g}(\mathbf{b}_i) = -\mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i / 2$. The log-likelihood contribution for the $i$th subject can be written as:

$$\begin{aligned} f(\mathbf{y}_i) &= \log\left[\frac{\lambda \cdot \rho \cdot (\prod_{j=1}^{n_i} y_{ij}^{\rho-1})}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot \int \exp\left[\tilde{f}(\mathbf{y}_i) + \tilde{g}(\mathbf{b}_i)\right] \cdot d\mathbf{b}_i\right] \\ &\propto -\frac{1}{2} \cdot \log \mid D \mid + \log K_i, \end{aligned} \tag{F.28}$$

where $K_i = \int I_i \cdot d\mathbf{b}_i = c \cdot \tilde{\phi}(\mathbf{y}_i)$ and $I_i = \exp\left[\tilde{f}(\mathbf{y}_i) + \tilde{g}(\mathbf{b}_i)\right]$.

The first derivative of the log-likelihood with respect to the fixed effects takes the following form:

$$\begin{aligned} \frac{\partial l_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} &= \frac{1}{K_i} \cdot \int I_i \cdot \left\{\sum_{j=1}^{n_i} \left[\mathbf{x}_{ij} - \lambda \cdot y_{ij}^{\rho} \cdot \exp(\boldsymbol{\mu}_{ij}) \cdot \mathbf{x}_{ij}\right]\right\} \cdot d\mathbf{b}_i \\ &= \sum_{j=1}^{n_i} \mathbf{x}_{ij} - \frac{1}{K_i} \cdot \sum_{j=1}^{n_i} \left[\lambda \cdot y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \int I_i \cdot \exp(\boldsymbol{\mu}_{ij}) \cdot d\mathbf{b}_i\right]. \end{aligned} \tag{F.29}$$

The component relative to the integral part in ($F$.29) can be rewritten as:

$$\begin{aligned} \int I_i \cdot \exp(\boldsymbol{\mu}_{ij}) \cdot d\mathbf{b}_i &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \int \exp(\boldsymbol{\mu}_{ij}) \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot d\mathbf{b}_i \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \exp(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}) \cdot \int \exp(\mathbf{z}_i^{'} \cdot \mathbf{b}_i) \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \cdot d\mathbf{b}_i \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \exp(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}) \cdot \exp\left(\frac{1}{2} \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij}\right) \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \exp\left(\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \frac{1}{2} \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij}\right). \end{aligned} \tag{F.30}$$

Expanding the component functions in ($F$.29) leads to:

$$\frac{\partial l_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i} \mathbf{x}_{ij} - \lambda \cdot \sum_{j=1}^{n_i} y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \exp(\boldsymbol{\mu}_{ij}). \tag{F.31}$$

Further, the first derivative with respect to the $D$-components is:

$$\frac{\partial l_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot (D^{-1})_{jk} + \frac{1}{2 \cdot K_i} \cdot \int I_i \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \cdot d\mathbf{b}_i. \tag{F.32}$$

Solving the integral expression leads to:

$$\begin{aligned} \int I_i \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \cdot d\mathbf{b}_i &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \mathrm{E}(\mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i \mid \mathbf{y}_i) \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \mathrm{E}\left[\mathrm{tr}(\mathbf{b}_i^{'} \cdot D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathbf{b}_i)\right] \\ &= c \cdot \tilde{\phi}(\mathbf{y}_i) \cdot \mathrm{tr}\left[D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i \mid \mathbf{y}_i)\right], \end{aligned} \tag{F.33}$$

where $E_{jk}$ defined in Section $F$.1.2. Expectation $\mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i \mid \mathbf{y}_i)$ is derived using the closed form for the Weibull-normal model:

$$\begin{aligned} \mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i \mid \mathbf{y}_i) &= \int \mathbf{b}_i^{'} \cdot \mathbf{b}_i \cdot \tilde{\phi}(\mathbf{b}_i \mid \mathbf{y}_i) \\ &= \int \mathbf{b}_i^{'} \cdot \mathbf{b}_i \cdot \frac{\tilde{\phi}(\mathbf{y}_i \mid \mathbf{b}_i) \cdot \tilde{\eta}(\mathbf{b}_i)}{\tilde{\phi}(\mathbf{y}_i)} \cdot \mathbf{b}_i \\ &= \frac{1}{\tilde{\phi}(\mathbf{y}_i)} \cdot \int \mathbf{b}_i^{'} \cdot \mathbf{b}_i \cdot \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \\ &\quad \cdot \exp\left[(m_j + 1) \cdot \boldsymbol{\mu}_{ij} - \frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i\right] \cdot d\mathbf{b}_i \\ &= \frac{1}{\tilde{\phi}(\mathbf{y}_i)} \cdot \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \frac{1}{(2 \cdot \pi)^{q/2} \cdot \mid D \mid^{1/2}} \\ &\quad \cdot \int \mathbf{b}_i^{'} \cdot \mathbf{b}_i \cdot \exp\left[(m_j + 1) \cdot \boldsymbol{\mu}_{ij} - \frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i\right] \cdot d\mathbf{b}_i, \end{aligned} \tag{F.34}$$

where $\tilde{\phi}(\mathbf{y}_i) = K_i/c$. Reorganizing the components of the exponential expression in the integrand of ($F$.34) leads to:

$$-\frac{1}{2} \cdot \mathbf{b}_i^{'} \cdot D^{-1} \cdot \mathbf{b}_i + (m_j + 1) \cdot \boldsymbol{\mu}_{ij} = -\frac{1}{2} \cdot (\mathbf{b}_i - \mathbf{k})^{'} \cdot D^{-1} \cdot (\mathbf{b}_i - \mathbf{k}) + \mathbf{l}, \tag{F.35}$$

with

$$\mathbf{k} = (m_j + 1) \cdot D \cdot \mathbf{z}_{ij}, \;\; \mathbf{l} = (m_j + 1) \cdot \left[\mathbf{x}_{ij}^{'} \cdot \boldsymbol{\beta} + \frac{1}{2} \cdot (m_j + 1) \cdot \mathbf{z}_{ij}^{'} \cdot D \cdot \mathbf{z}_{ij}\right], \;\; \tilde{\mathbf{b}}_i = \mathbf{b}_i - \mathbf{k}. \tag{F.36}$$

Rewriting $\tilde{\phi}(\mathbf{y}_i)$ leads to:

$$\tilde{\phi}(\mathbf{y}_i) = \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \exp(\mathbf{l}). \tag{F.37}$$

Combining $(F.34)$ and $(F.35)$ produces:

$$\begin{aligned}
\mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i \mid \mathbf{y}_i) &= \frac{1}{\tilde{\phi}(\mathbf{y}_i)} \cdot \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \frac{\exp(\mathbf{l})}{(2\cdot\pi)^{q/2}\cdot \mid D \mid^{1/2}} \\
&\quad \cdot \int (\tilde{\mathbf{b}}_i + \mathbf{k})^{'} \cdot (\tilde{\mathbf{b}}_i + \mathbf{k}) \cdot \exp\left(-\frac{1}{2} \cdot \tilde{\mathbf{b}}_i^{'} \cdot D^{-1} \cdot \tilde{\mathbf{b}}_i\right) \cdot d\tilde{\mathbf{b}}_i \\
&= \frac{1}{\tilde{\phi}(\mathbf{y}_i)} \cdot \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \exp(\mathbf{l}) \cdot \mathrm{E}\left[(\tilde{\mathbf{b}}_i + \mathbf{k})^{'} \cdot (\tilde{\mathbf{b}}_i + \mathbf{k})\right] \\
&= \frac{1}{\tilde{\phi}(\mathbf{y}_i)} \cdot \sum_{\mathbf{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \cdot \lambda^{m_j+1} \cdot \rho \cdot y_{ij}^{(m_j+1)\cdot\rho-1} \cdot \exp(\mathbf{l}) \cdot \mathrm{E}(\mathbf{b}_i^{'} \cdot \mathbf{b}_i) \\
&= \mathrm{Var}(\mathbf{b}_i).
\end{aligned} \tag{F.38}$$

Plugging $(F.37)$ into $(F.33)$ and $(F.32)$ yields:

$$\begin{aligned}
\frac{\partial l_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} &= -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot (D^{-1})_{jk} + \frac{1}{2} \cdot \mathrm{tr}\left[D^{-1} \cdot E_{jk} \cdot D^{-1} \cdot \mathrm{Var}(\mathbf{b}_i)\right] \\
&= -\frac{1}{2} \cdot (2 - \delta_{jk}) \cdot \left[(D^{-1})_{jk} - (D^{-1} \cdot D^{-1})_{jk} \cdot \mathrm{Var}(\mathbf{b}_i)\right]
\end{aligned} \tag{F.39}$$

where $\delta_{jk}$ is as before.

The vector $\boldsymbol{\Delta}_i$ of first-order partial derivative of the contribution of the $i$th subject to the log-likelihood is now given by:

$$\boldsymbol{\Delta}_i = \begin{bmatrix}
\sum_{j=1}^{n_i} \mathbf{x}_{ij} - \lambda \cdot \sum_{j=1}^{n_i} y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \exp(\boldsymbol{\mu}_{ij}) \\
-\frac{1}{2} \cdot (D^{-1})_{11} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{11} \cdot \mathrm{Var}(\mathbf{b}_i) \\
-(D^{-1})_{12} + (D^{-1} \cdot D^{-1})_{12} \cdot \mathrm{Var}(\mathbf{b}_i) \\
-\frac{1}{2} \cdot (D^{-1})_{22} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{22} \cdot \mathrm{Var}(\mathbf{b}_i) \\
\vdots \\
-(D^{-1})_{q-1,q} + (D^{-1} \cdot D^{-1})_{q-1,q} \cdot \mathrm{Var}(\mathbf{b}_i) \\
-\frac{1}{2} \cdot (D^{-1})_{qq} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{qq} \cdot \mathrm{Var}(\mathbf{b}_i)
\end{bmatrix} \tag{F.40}$$

Rewriting $\|\boldsymbol{\Delta}_i\|$ as the sum of squares of the contributions for the $i$th individual yields:

$$\begin{aligned}
\|\boldsymbol{\Delta}_i\| &= \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} - \mathbf{Q}_i\right)^2 + \sum_{k=1}^{q} \left[-\frac{1}{2} \cdot (D^{-1})_{kk} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{kk} \cdot \mathrm{Var}(\mathbf{b}_i)\right]^2 \\
&\quad + \sum_{k<l} \left[-(D^{-1})_{kl} + (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\right]^2 \\
&= \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij}\right) \cdot \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij}\right)^{'} - 2 \cdot \sum_{j=1}^{n_i} \mathbf{x}_{ij} \cdot \mathbf{Q}_i^{'} + \mathbf{Q}_i \cdot \mathbf{Q}_i^{'} \\
&\quad + \sum_{k,l} \left[-\frac{1}{2} \cdot (D^{-1})_{kl} + \frac{1}{2} \cdot (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\right]^2,
\end{aligned} \tag{F.41}$$

where $\mathbf{Q}_i = \lambda \cdot \sum_{j=1}^{n_i} y_{ij}^{\rho} \cdot \mathbf{x}_{ij} \cdot \exp(\boldsymbol{\mu}_{ij})$. Write $C_i = C_{1i} + C_{2i}$, with:

$$\begin{aligned} C_{1i} &= 2 \cdot \|\ddot{L}^{-1}\| \cdot \left(\|\mathbf{x}_i\|^2 - 2 \cdot \mathbf{x}_i \cdot \mathbf{Q}_i + \|\mathbf{Q}_i\|^2\right) \cdot \cos(\varphi_i), \\ C_{2i} &= \frac{1}{2} \cdot \|\ddot{L}^{-1}\| \cdot \|(D^{-1})_{kl} - (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\|^2 \cdot \cos(\varphi_i), \end{aligned} \tag{F.42}$$

where $\mathbf{x}_i = \sum_{j=1}^{n_i} \mathbf{x}_{ij}$. Note that $C_{1i}$ and $C_{2i}$ are the contributions of the $i$th subject to local influence contributions $C_i$ from $\boldsymbol{\beta}$ and $D$, respectively. Rewriting the component of $C_{2i}$ leads to:

$$\begin{aligned} \|(D^{-1})_{kl} - (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\|^2 &= \mathrm{tr}\left\{\left[(D^{-1})_{kl} - (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\right]^2\right\} \\ &= \mathrm{tr}\left[(D^{-1})_{kl}^2\right] - \mathrm{tr}\left[2 \cdot (D^{-1})_{kl} \cdot (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\right] \\ &\quad + \mathrm{tr}\left[(D^{-1} \cdot D^{-1})_{kl}^2 \cdot \mathrm{Var}(\mathbf{b}_i)^2\right]. \end{aligned} \tag{F.43}$$

It then follows that:

$$\begin{aligned} C_{1i} &= 2 \cdot \|\ddot{L}^{-1}\| \cdot \left(\|\mathbf{x}_i\|^2 - 2 \cdot \mathbf{x}_i \cdot \mathbf{Q}_i + \|\mathbf{Q}_i\|^2\right) \cdot \cos(\varphi_i), \\ C_{2i} &= \frac{1}{2} \cdot \|\ddot{L}^{-1}\| \cdot \cos(\varphi_i) \cdot \mathrm{tr}\left[(D^{-1})_{kl}^2\right] - \mathrm{tr}\left[2 \cdot (D^{-1})_{kl} \cdot (D^{-1} \cdot D^{-1})_{kl} \cdot \mathrm{Var}(\mathbf{b}_i)\right] \\ &\quad + \mathrm{tr}\left[(D^{-1} \cdot D^{-1})_{kl}^2 \cdot \mathrm{Var}(\mathbf{b}_i)^2\right]. \end{aligned} \tag{F.44}$$

Hence, the interpretable components of $C_i$ for the Weibull normal model can be described using the length of fixed effect ($\|\mathbf{x}_i\|^2$) and the squared of random effect variability ($\mathrm{Var}(\mathbf{b}_i)^2$), in analogy with the Poisson-normal model.

# Chapter G

# Supplementary Material for Chapter 11

## G.1 Descriptive Statistics of the Asthma dataset

Before applying specific modeling techniques to the asthma dataset, descriptive analysis is advisable a priori. Therefore, descriptive statistics of all variables are done, leading to the following main conclusions:

1. An average of 7.66 follow-ups were present per patient;
2. From these follow-ups, an average of 6.71 asthma attacks was present per patient, where patients with ID 74, 107 and 177 have the most attacks, i.e., 23, 22 and 21, respectively, while ID's 69, 89 and 228 didn't had any attacks at all;
3. From the patients that were present in the study, 93.53% had a censored observation (due to drop-out, lost to follow-up, etc.);
4. A total of 51.72% patients received the drug, while the other 48.28% received the placebo;

In Chapter 11, main interest lies in the development of an appropriate modeling fit for the time to recurrence of an asthma attack, i.e., a repeated time-to-event outcome. Since Weibull and Exponential are both popular tools, simple diagnostics are first done on the distribution of this outcome. From Figure G.1, both choices seems reasonable. Here, the exponential one is chosen as underlying theoretical distribution.

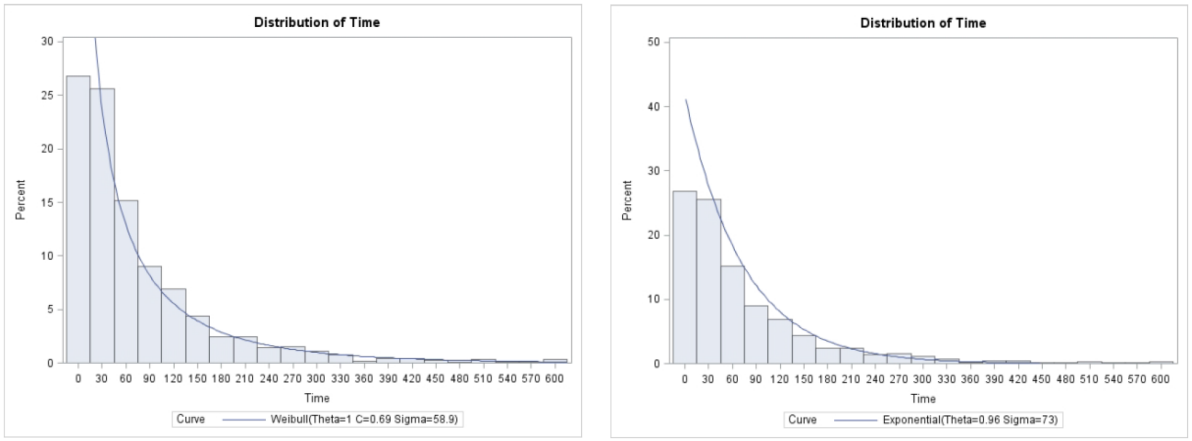


**Figure G.1:** *Graphical diagnostics check of the time to recurrence of an asthma attack for Weibull (left) and Exponential (right) distribution.*

## G.2 Relevant SAS code for analyzing the Asthma dataset

```
/*****************************************************************
SOFTWARE: SAS 9.3.
OBJECTIVE: A full analysis of the Asthma dataset using the WGN
model (4.1)-(4.3), with rho equal to 1, i.e., the exponential-
gamma-normal model;
DATASET: Example 9 of Duchateau & Janssen (2008);
VARIABLE DESCRIPTION:
Patid: Patient ID;
Begin and End: time interval between events for each patient;
Status: Right censoring indicator (1 = Asthma Attack, 0 = Censo-
red);
Drug: Treatment indicator (1 = Drug, 0 = Placebo).
REFERENCES: Iddi et al (2014) & Molenberghs et al (2013).

Thanks given to Samuel Iddi, Achmad Efendi, Trias Wahyuni Rakhma-
wati and Geert Molenberghs, for providing relevant SAS code.
*****************************************************************/
libname asthma 'D:\Mijn Documenten\Master Statistiek\Masterthesis
                \Data';

data asthma;
        set asthma.asthma;
run;
```

### G.2.1 Several Hierarchical Modeling Strategies with different Estimation Strategies

```
/* Exponential model, via NLMIXED */
```

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

/* Exponential-Gamma model */
proc nlmixed data=asthma tech=quanew;
        parms Beta_0=-3 Beta_1=-0.2 lambda=1 alpha=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1);
        expeta = exp(eta);
        ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)
             + (rho-1)*log(Time) + eta
             - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        model Time ~ general(ll);
run;

/* Exponential-Normal model */
proc nlmixed data=asthma qpoints=50;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        ll = log(rho) + log(gamma) + (gamma-1)*log(Time)
             + eta - rho*(Time**gamma)*expeta;
        model Time ~ general(ll);
        random b1 ~ normal(0,sigma**2) subject=Patid;
run;

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

```
/* Combined model without censoring */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds lambda > 0, alpha > 0;
        parms Beta_0=-3 Beta_1=-0.2 lambda = 1 alpha=3.3
              sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
run;

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

/* Pseudo-likelihood, No censoring */
data asma1;
        set work.asthma;
        keep patid drug time;
run;

proc sort data=asma1;
        by patid;
run;

%macro pseudosurv(data= ,subject= ,fixed= ,response= );
        proc freq data=&data;
                tables &subject / out=freq noprint;
        run;
        data freq;
                set freq;
```

```
        keep &subject count;
run;
proc iml;
        use freq;
        read all into f;
        use &data;
        read all into y[colname=coln];
        nf=nrow(f);
        ny=nrow(y);
        do i=1 to nf;
                codesubject = f[i,1];
                free newdata;
                do b=1 to ny;
                        if y[b,1]=codesubject then do;
                                m=y[b,];
                                newdata=newdata//m;
                        end;
                end;
                ni=nrow(newdata);
                npair=ni*(ni-1)/2;
                pairs=J(npair,4,0);
                h=1;
                do j=1 to (ni-1);
                        do k=j+1 to ni;
                                pairs[h,1]=codesubject;
                                pairs[h,2]=newdata[1,2];
                                pairs[h,3]=newdata[j,3];
                                pairs[h,4]=newdata[k,3];
                                h=h+1;
                        end;
                end;
        result=result//pairs;
        end;
        x=result;
        x1=x[,1]||x[,2];
        x2=x[,3]||x[,4];
        npair=nrow(x1);
        z1=x1//x1;
        call sort(z1,{1 2});
        z2=shape(x2,npair*2,1);
        a1=1:npair;
        a2=a1'||a1';
        z3=shape(a2,npair*2,1);
        z=z1||z2||z3;
        create last from z; append from z;
quit;
proc freq data=last;
```

```
                tables col1 / out=freq noprint;
        run;
        data freq;
                set freq;
                keep col1 count;
        run;
        proc iml;
                use freq;
                read all into f;
                use last;
                read all into w;
                nf=nrow(f);
                do i=1 to nf;
                        a1=1:f[i,2]/2;
                        a2=a1//a1;
                        a3=shape(a2`,f[i,2],1);
                        b=b//a3;
                end;
                c=w||b;
                cname = {"Subj", "Fixed", "Resp", "Pairnum",
                         "Pairsubj"};
                create last1 from c[colname=cname]; append from c;
        quit;
%mend pseudosurv;

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

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

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

/* Exponential model via NLMIXED */
proc nlmixed data=asma2 tech=newrap;
        parms Beta_0=-3.9 Beta_1=-0.09 gamma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1);
        expeta = exp(eta);
        ll = log(rho) + log(gamma) + (gamma-1)*log(Time)
```

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

/*************************************************************
Exponential model with gamma frailty and random Effects,
fitting model that has five parameters in it
*************************************************************/
proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
        bounds lambda > 0, alpha > 0;
        parms Beta_0=-3.3 Beta_1=-0.08 lambda=1 alpha=3.3
              sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=pairnum;
        ods output ParameterEstimates=initial;
run;

data initial;
        set initial;
        keep estimate;
run;

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

data asma3;
        set asma2;
        drop pairnum;
run;
```

```
proc sort data=asma3;
        by patid;
run;

%macro se5;
        %do i=1 %to 230;
                data asm&i;
                        set asma3;
                        where patid=&i;
                run;
                proc sort data=asm&i;
                        by pairsubj;
                run;
                proc nlmixed data=asm&i tech=quanew qpoints=50
                                noad maxit=1000 start hess;
                        bounds lambda > 0, alpha > 0;
                        parms Beta_0=&es0 Beta_1=&es1
                              lambda=&es2 alpha=&es3 sigma=&es4;
                        rho=1;
                        eta = Beta_0 + Beta_1*(Drug=1) + b1;
                        expeta = exp(eta);
                        ll = log(lambda) + log(rho)
                              + (alpha+1)*log(alpha)
                              + (rho-1)*log(Time) + eta
                              - (alpha+1)*log(lambda*(Time**rho)
                              *expeta + alpha);
                        model Time ~ general(ll);
                        random b1 ~ normal(0, sigma**2)
                               subject=pairsubj;
                        ods output startingvalues=grad
                                   startinghessian=hes&i;
                run;
                data grad&i;
                        set grad;
                        keep gradient;
                run;
        %end;
        data score;
                set _null_;
        run;
        data hess;
                set _null_;
        run;
        %do i=1 %to 230;
                data score;
                        set score grad&i;
```

```
        run;
        data hess;
                set hess hes&i;
        run;
%end;
data hess;
        set hess;
        drop row parameter;
run;
proc iml;
        use score;
        read all into s;
        use hess;
        read all into h;
        nid=nrow(s)/5;
        nidh=nrow(h)/5;
        a=1:nid;
        b=J(5,1,1);
        c=b@(a`);
        a1=1:nidh;
        b1=J(5,1,1);
        c1=b1@(a1`);
        create s1 from c;
        append from c;
        create s2 from c1;
        append from c1;
quit;
proc sort data=s1;
        by col1;
run;
proc sort data=s2;
        by col1;
run;
proc iml;
        use score;
        read all into scor;
        use hess;
        read all into hes;
        use s1;
        read all into j1;
        use s2;
        read all into j2;
        scor1=j1||scor;
        hes1=j2||hes;
        I1=J(5,5,0);
        do i=1 to 230;
                free g;
```

```
                        do b=1 to (230*5);
                                if scor1[b,1]=i then do;
                                        m=scor1[b,2];
                                        g=g//m;
                                end;
                        end;
                        sg=g*g';
                        I1=I1+sg;
                end;
                I0=J(5,5,0);
                do i=1 to 226;
                        free g1;
                        do b=1 to (226*5);
                                if hes1[b,1]=i then do;
                                        m=hes1[b,2]||hes1[b,3]
                                           ||hes1[b,4]||hes1[b,5]
                                           ||hes1[b,6];
                                        g1=g1//m;
                                end;
                        end;
                        I0=I0+g1;
                end;
                vmatn=inv(I0);
                naiv_se=sqrt(vecdiag(vmatn));
                vmat=inv(I0)*I1*inv(I0);
                emp_se=sqrt(vecdiag(vmat));
                /*naive and empr se*/
                se=naiv_se||emp_se;
                print se;
        quit;
%mend;
%se5;

/*************************************************************
Weibull model with gamma frailty and random Effects,
fitting model that has four parameters in it, lambda fixed
*************************************************************/
proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
        bounds alpha > 0;
        parms Beta_0=-3.3 Beta_1=-0.08 alpha=3.3 sigma=1;
        rho=1;
        lambda=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)
             + (rho-1)*log(Time) + eta
             - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
```

```
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=pairnum;
        ods output ParameterEstimates=initial2;
run;

data initial2;
        set initial2;
        keep estimate;
run;

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

proc sort data=asma3;
        by patid;
run;
%macro se4;
        %do i=1 %to 230;
                data asm&i;
                        set asma3;
                        where patid=&i;
                run;
                proc sort data=asm&i;
                        by pairsubj;
                run;
                proc nlmixed data=asm&i tech=quanew qpoints=50
                                maxit=1000 start hess;
                        bounds alpha > 0;
                        parms Beta_0=&esa0 Beta_1=&esa1
                              alpha=&esa2 sigma=&esa3;
                        rho=1;
                        lambda=1;
                        eta = Beta_0 + Beta_1*(Drug=1) + b1;
                        expeta = exp(eta);
                        ll = log(lambda) + log(rho)
                             + (alpha+1)*log(alpha)
                             + (rho-1)*log(Time) + eta
```

```
                        - (alpha+1)*log(lambda*(Time**rho)
                        *expeta + alpha);
                model Time ~ general(ll);
                random b1 ~ normal(0, sigma**2)
                        subject=pairsubj;
                ods output startingvalues=grad
                                startinghessian=hes&i;
        run;
        data grad&i;
                set grad;
                keep gradient;
        run;
%end;
data score;
        set _null_;
run;
data hess;
        set _null_;
run;
%do i=1 %to 230;
        data score;
                set score grad&i;
        run;
        data hess;
                set hess hes&i;
        run;
%end;
data hess;
        set hess;
        drop row parameter;
run;
proc iml;
        use score;
        read all into s;
        use hess;
        read all into h;
        nid=nrow(s)/4;
        nidh=nrow(h)/4;
        a=1:nid;
        b=J(4,1,1);
        c=b@(a');
        a1=1:nidh;
        b1=J(4,1,1);
        c1=b1@(a1');
        create s1 from c;
        append from c;
        create s2 from c1;
```

```
        append from c1;
quit;
proc sort data=s1;
        by col1;
run;
proc sort data=s2;
        by col1;
run;
proc iml;
        use score;
        read all into scor;
        use hess;
        read all into hes;
        use s1;
        read all into j1;
        use s2;
        read all into j2;
        scor1=j1||scor;
        hes1=j2||hes;
        I1=J(4,4,0);
        do i=1 to 230;
                free g;
                do b=1 to (230*4);
                        if scor1[b,1]=i then do;
                                m=scor1[b,2];
                                g=g//m;
                        end;
                end;
                sg=g*g';
                I1=I1+sg;
        end;
        I0=J(4,4,0);
        do i=1 to 226;
                free g1;
                do b=1 to (226*4);
                        if hes1[b,1]=i then do;
                                m=hes1[b,2]||hes1[b,3]
                                   ||hes1[b,4]||hes1[b,5];
                                g1=g1//m;
                        end;
                end;
                I0=I0+g1;
        end;
        vmatn=inv(I0);
        naiv_se=sqrt(vecdiag(vmatn));
        vmat=inv(I0)*I1*inv(I0);
        emp_se=sqrt(vecdiag(vmat));
```

```
                se=naiv_se||emp_se;
                print se;
        quit;
%mend;
%se4;

/* Pseudo-likelihood, Censoring */
%macro plcensor(data= ,subject= ,status= status, fixed= ,
                response= );
        proc freq data=&data;
                tables &subject / out=freq noprint;
        run;
        data freq;
                set freq;
                keep &subject count;
        run;
        proc iml;
                use freq;
                read all into f;
                use &data;
                read all into y[colname=coln];
                nf=nrow(f);
                ny=nrow(y);
                do i=1 to nf;
                        codesubject = f[i,1];
                        free newdata;
                        do b=1 to ny;
                                if y[b,1]=codesubject then do;
                                        m=y[b,];
                                        newdata=newdata//m;
                                end;
                        end;
                        ni=nrow(newdata);
                        Npair=ni*(ni-1)/2;
                        pairs=J(Npair,6,0);
                        h=1;
                        do j=1 to (ni-1);
                                do k=j+1 to ni;
                                        pairs[h,1]=codesubject;
                                        pairs[h,2]=newdata[j,2];
                                        pairs[h,3]=newdata[k,2];
                                        pairs[h,4]=newdata[1,3];
                                        pairs[h,5]=newdata[j,4];
                                        pairs[h,6]=newdata[k,4];
                                        h=h+1;
                                end;
                        end;
```

```
                result=result//pairs;
            end;
            x=result;
            x1=x[,1]||x[,4];
            x2=x[,2]||x[,3];
            x3=x[,5]||x[,6];
            z1=x1//x1;
            call sort(z1,{1 2});
            z2=shape(x2,nrow(x2)*2,1);
            z3=shape(x3,nrow(x3)*2,1);
            npair=nrow(z1)/2;
            a1=1:npair;
            a2=a1//a1;
            z4=shape(a2 ,nrow(z1),1);
            z=z1||z2||z3||z4;
            create last from z;
            append from z;
    quit;
    proc freq data=last;
            tables col1 / out=freq noprint;
    run;
    data freq;
            set freq;
            keep col1 count;
    run;
    proc iml;
            use freq;
            read all into f;
            use last;
            read all into w[colname=coln];
            nf=nrow(f);
            do i=1 to nf;
                    a1=1:f[i,2]/2;
                    a2=a1//a1;
                    a3=shape(a2`,f[i,2],1);
                    b=b//a3;
            end;
            c=w||b;
            cname = {"Subj", "Fixed", "Status", "Resp",
                     "Pairnum", "Pairsubj"};
            create last1 from c[colname=cname];append from c;
    quit;
%mend plcensor;

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

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

proc sort data=asma2;
        by patid;
run;

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

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

/*************************************************************
Weibull model with gamma frailty and random Effects,
incorporating censoring, fitting model that has five parameters
in it.
*************************************************************/
proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
        bounds lambda > 0, alpha > 0;
        parms Beta_0=-3.8 Beta_1=-0.15 lambda=1 alpha=5 sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Pairnum;
        ods output parameterestimates=initial;
run;

data initial;
        set initial;
        keep estimate;
run;
```

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

data asma3;
        set asma2;
        drop pairnum;
run;

proc sort data=asma3;
        by patid;
run;

%macro se5;
        %do i=1 %to 230;
                data asm&i;
                        set asma3;
                        where patid=&i;
                run;
                proc sort data=asm&i;
                        by pairsubj;
                run;
                proc nlmixed data=asm&i tech=quanew qpoints=50
                                noad maxit=1000 start hess;
                        bounds lambda > 0, alpha > 0;
                        parms Beta_0=&es0 Beta_1=&es1 lambda=&es2
                              alpha=&es3 sigma=&es4;
                        rho=1;
                        eta = Beta_0 + Beta_1*(Drug=1) + b1;
                        expeta = exp(eta);
                        c0 = 1/((1 + lambda*expeta*(Time**rho)
                             *(1/alpha))**alpha);
                        c1 = log(lambda) + log(rho)
                              + (alpha+1)*log(alpha)
                              + (rho-1)*log(Time) + eta
```

```
                        - (alpha+1)*log(lambda*(Time**rho)
                        *expeta + alpha);
                ll = (status=0)*log(c0) + (status=1)*c1;
                model Time ~ general(ll);
                random b1 ~ normal(0, sigma**2)
                        subject=pairsubj;
                ods output startingvalues=grad
                        startinghessian=hes&i;
        run;
        data grad&i;
                set grad;
                keep gradient;
        run;
%end;
data score;
        set _null_;
run;
data hess;
        set _null_;
run;
%do i=1 %to 230;
        data score;
                set score grad&i;
        run;
        data hess;
                set hess hes&i;
        run;
%end;
data hess;
        set hess;
        drop row parameter;
run;
proc iml;
        use score;
        read all into s;
        use hess;
        read all into h;
        nid=nrow(s)/5;
        nidh=nrow(h)/5;
        a=1:nid;
        b=J(5,1,1);
        c=b@(a');
        a1=1:nidh;
        b1=J(5,1,1);
        c1=b1@(a1');
        create s1 from c;
        append from c;
```

```
        create s2 from c1;
        append from c1;
quit;
proc sort data=s1;
        by col1;
run;
proc sort data=s2;
        by col1;
run;
proc iml;
        use score;
        read all into scor;
        use hess;
        read all into hes;
        use s1;
        read all into j1;
        use s2;
        read all into j2;
        scor1=j1||scor;
        hes1=j2||hes;
        I1=J(5,5,0);
        do i=1 to 230;
                free g;
                do b=1 to (230*5);
                        if scor1[b,1]=i then do;
                                m=scor1[b,2];
                                g=g//m;
                        end;
                end;
                sg=g*g';
                I1=I1+sg;
        end;
        I0=J(5,5,0);
        do i=1 to 226;
                free g1;
                do b=1 to (226*5);
                        if hes1[b,1]=i then do;
                                m=hes1[b,2]||hes1[b,3]
                                    ||hes1[b,4]||hes1[b,5]||hes1[b
                                g1=g1//m;
                        end;
                end;
        I0=I0+g1;
        end;
        vmatn=inv(I0);
        naiv_se=sqrt(vecdiag(vmatn));
        vmat=inv(I0)*I1*inv(I0);
```

```
                emp_se=sqrt(vecdiag(vmat));
                se=naiv_se||emp_se;
                print se;
        quit;
%mend;
%se5;

/*************************************************************
Weibull model with gamma frailty and random Effects,
incorporating censoring, fitting model that has four parameters
in it, lambda fixed.
*************************************************************/
proc nlmixed data=asma2 tech=quanew qpoints=50 maxit=1000;
        bounds alpha > 0;
        parms Beta_0=-3.8 Beta_1=-0.1 alpha=4.5 sigma=0.2;
        rho=1;
        lambda=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))**alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Pairnum;
        ods output parameterestimates=initial2;
run;

data initial2;
        set initial2;
        keep estimate;
run;

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

```
proc sort data=asma3;
        by patid;
run;

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

```
        run;
        data hess;
                set hess hes&i;
        run;
%end;
data hess;
        set hess;
        drop row parameter;
run;
proc iml;
        use score;
        read all into s;
        use hess;
        read all into h;
        nid=nrow(s)/4;
        nidh=nrow(h)/4;
        a=1:nid;
        b=J(4,1,1);
        c=b@(a`);
        a1=1:nidh;
        b1=J(4,1,1);
        c1=b1@(a1`);
        create s1 from c;
        append from c;
        create s2 from c1;
        append from c1;
quit;
proc sort data=s1;
        by col1;
run;
proc sort data=s2;
        by col1;
run;
proc iml;
        use score;
        read all into scor;
        use hess;
        read all into hes;
        use s1;
        read all into j1;
        use s2;
        read all into j2;
        scor1=j1||scor;
        hes1=j2||hes;
        I1=J(4,4,0);
        do i=1 to 230;
        free g;
```

```
                        do b=1 to (230*4);
                                if scor1[b,1]=i then do;
                                        m=scor1[b,2];
                                        g=g//m;
                                end;
                        end;
                        sg=g*g';
                        I1=I1+sg;
                end;
                I0=J(4,4,0);
                do i=1 to 226;
                free g1;
                        do b=1 to (226*4);
                                if hes1[b,1]=i then do;
                                        m=hes1[b,2]||hes1[b,3]
                                           ||hes1[b,4]||hes1[b,5];
                                        g1=g1//m;
                                end;
                        end;
                        I0=I0+g1;
                end;
                vmatn=inv(I0);
                naiv_se=sqrt(vecdiag(vmatn));
                vmat=inv(I0)*I1*inv(I0);
                emp_se=sqrt(vecdiag(vmat));
                se=naiv_se||emp_se;
                print se;
        quit;
%mend;
%se4;

/* Special case: Combined model without censoring and setting
   lambda=1 */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds alpha > 0;
        parms Beta_0=-3 Beta_1=-0.2 alpha=3.3 sigma=1;
        rho=1;
        lambda = 1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)
             + (rho-1)*log(Time) + eta
             - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
run;
```

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

### G.2.2 The Combined Model and its Marginalized Version

```
/* Hierarchical Exponential-Gamma-Normal with censoring */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds lambda>0, alpha>0;
        parms Beta1=-0.08 lambda=1 alpha=3.3 sigma1=1 sigma2=1;
        rho=1;
        eta=(Beta1+b2)*(Drug=1)+b1;
        expeta=exp(eta);
        c0=1/((1+lambda*expeta*(Time**rho)*(1/alpha))**alpha);
        c1=log(lambda)+log(rho)+(alpha+1)*log(alpha)
            + (rho-1)*log(Time) +eta
            - (alpha+1)*log(lambda*(Time**rho)*expeta+alpha);
        ll=(status=0)*log(c0)+(status=1)*c1;
        model Time~general(ll);
        random b1 b2 ~normal([0,0],[sigma1**2,0,sigma2**2])
               subject=Patid;
        estimate "Var of R.E1.s" sigma1**2;
        estimate "Var of R.E2.s" sigma2**2;
run;

/* Marginalized Exponential-Gamma-Normal with censoring */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds lambda>0, alpha>0;
        parms Beta1=-0.08 lambda=1 alpha=3.3 sigma1=1 sigma2=1;
        rho=1;
        eta=(Beta1+b2)*(Drug=1);
        delta=eta-(sigma1*sigma1+sigma2*sigma2*(Drug=1)
              *(Drug=1))/2;
```

```
        etas=delta+b1;
        expeta=exp(etas);
        c0=1/((1+lambda*expeta*(Time**rho)*(1/alpha))**alpha);
        c1=log(lambda)+log(rho)
            +(alpha+1)*log(alpha)+(rho-1)*log(Time) +eta
            - (alpha+1)*log(lambda*(Time**rho)*expeta+alpha);
        ll=(status=0)*log(c0)+(status=1)*c1;
        model Time~general(ll);
        random b1 b2 ~normal([0,0],[sigma1**2,0,sigma2**2])
              subject=Patid;
        estimate "Var of R.E1.s" sigma1**2;
        estimate "Var of R.E2.s" sigma2**2;
run;
```

### G.2.3 Gradient Function for Assessing Fit of Random-Effects Distribution

```
/* Plot EB estimates normal random effects */
proc sgplot data=EB_out;
        histogram Estimate;
        density Estimate;
        density Estimate / type=kernel;
run;

/* Plot EB estimates gamma random effects */
/*Integration over theta_ij (Iddi et al, 2014)*/
proc iml;
        use work.asthma;
        read all into data;
        M = nrow(data);

        start fxn1(theta) global(alpha, lambda, rho, subdata, i,
                                 j, b0, bi);
                Drug = subdata[, 5];
                Time = subdata[, 7];
                term1 = b0[1] + b0[2]#Drug[j] + bi[i];
                kappa = theta#exp(term1);
                fy = lambda#rho#kappa#Time[j]##(rho-1)
                     #exp(-lambda#Time[j]##rho#kappa);
                term = (1/alpha)##alpha#exp(lgamma(alpha));
                ftheta = (1/term)#theta##(alpha-1)
                         #exp(-theta/(1/alpha));
                int = fy#ftheta;
        return(int);
        finish;

        start fxn2(theta) global(alpha, lambda, rho, subdata, i,
```

```
                        j, b0, bi);
        num = fxn1(theta);
        eps = 1E-10;
        lim2 = {0 .P};
        call quad(den, "fxn1", lim2) eps=eps scale=1
                cycles=8 msg="no";
        obj = (theta#num)/den;
return(obj);
finish;

start fxn3(theta) global(alpha, lambda, rho, subdata, i,
                        j, b0, bi, xres);
        num = fxn1(theta);
        eps = 1E-10;
        lim2 = {0 .P};
        call quad(den, "fxn1", lim2) eps=eps scale=1
                cycles=8 msg="no";
        obj = ((theta-xres)##2)#num/den;
return(obj);
finish;

/*Intials*/
N = 232;
use EB_out;
read all into bi;
bi = bi[, 3];
ebtheta = repeat(., M);
ebstd = repeat(., M);
b0 = {-4.0195 -0.1115 3.5633 0.3158 0.7882 1};
p = 0;
alpha = b0[3];
d = b0[4];
lambda = b0[5];
rho = b0[6];
do i = 1 to N;
        index1=t(loc(data[, 1]=i));
        subdata = data[index1, ];
        ni = nrow(subdata);
                do j = 1 to ni;
                        p = p+1;
                        eps = 1E-10;
                        lim = {0 .P};
                        call quad(xres, "fxn2", lim)
                                eps=eps scale=1 cycles=8
                                msg="no";
                        call quad(std, "fxn3", lim)
                                eps=eps scale=1 cycles=8
```

```
                                        msg="no";
                                    ebtheta[p] = xres;
                                    ebstd[p] = sqrt(std);
                                    *end;
                            end;
        end;
        create ebtheta from ebtheta;
        append from ebtheta;
        out = ebtheta||ebstd;
        create work.mythetacmwei from out[colname={'est' 'std'}];
        append from out;
quit;

/*Prediction*/
data allepi;
        merge work.asthma EB_out;
        by Patid;
run;
data allepi;
        merge allepi work.mythetacmwei;
run;

data work.AllResultWei;
        set allepi;
        predkappa = est*exp(-4.1993 - 0.0887*(Drug=1) + estimate);
        keep Patid Drug Time estimate StdErrPred est std
             predkappa;
run;

proc sql;
        create table work.theta as
                select distinct Patid as Patid, AVG(est)
                        as Empirical_Bayes_Estimate
        from work.allresultwei
        group by Patid
        order by Patid;
run;

proc sgplot data=theta;
  histogram Empirical_Bayes_estimate;
  density Empirical_Bayes_estimate / type=kernel;
run;

/* Calculation of log-likelihood values for fitted model, for
   each patient separately */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds lambda > 0, alpha > 0;
```

```
        parms Beta_0=-3 Beta_1=-0.11 lambda = 1 alpha=3.3
              sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))
             **alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
        by Patid;
        ods output parameters=out;
run;

data outmodel;
        set out;
        llmodel=-negloglike;
        keep Patid llmodel;
run;

/* Calculation of log-likelihood for grid of random-effect
   values, for each subject separately */
proc nlmixed data=asthma tech=quanew qpoints=50 maxit=1000;
        bounds lambda > 0, alpha > 0;
        parms Beta_0=-3 Beta_1=-0.11 lambda = 1 alpha=3.3
              sigma=1 b1=-5 to 5 by 0.1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))
             **alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        by Patid;
        ods output parameters=out1;
run;

data outgrid;
        set out1;
        llgrid=-negloglike;
        keep Patid b1 llgrid;
```

```
run;

/* Data management and calculation of gradient function
and pointwise confidence bands*/
data out_total;
        merge outgrid outmodel;
        by Patid;
run;

data out_total;
        set out_total;
        ratio=exp(llgrid-llmodel);
        keep Patid b1 ratio;
run;

proc sort data=out_total;
        by b1;
run;

proc means data=out_total alpha=0.05 mean var clm;
        var ratio;
        by b1;
        ods output summary=summary;
run;

data gradient;
        set summary;
        y=ratio_Mean;
        symbol="Gradient function";
        keep b1 y symbol;
        data lower;
                set summary;
                y=ratio_LCLM;
                symbol="95% CI - Lower";
                keep b1 y symbol;
        data upper;
                set summary;
                y=ratio_UCLM;
                symbol="95% CI - Upper";
                keep b1 y symbol;
        data figure;
        set gradient lower upper;
run;
```

### G.2.4 Local Influence Analysis

```
/*************************************************************
Local Influence for Exponential-Normal Model
*************************************************************/

data asthma_for_local;
        set asthma;
        count + 1;
        by Patid;
        if first.Patid then count = 1;
run;

data asthma_for_local;
        set asthma_for_local;
        int=1;
        placebo = (Drug=0);
        treatment = (Drug=1);
        ptime = placebo*count;
        ttime = treatment*count;
run;

proc freq data=asthma_for_local noprint;
        tables Patid /out=nfrec;
run;

proc sort data=asthma_for_local;
        by Patid;
run;

/* Exponential-normal model with censoring */
*use nlmixed the same slope hessian matrix for overall subject;
proc nlmixed data=asthma_for_local tech=quanew qpoints=50
                maxit=1000 hess start;
        bounds lambda > 0;
        parms Beta_0=-3 Beta_1=-0.11 lambda = 1 sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = log(rho)-lambda*(Time**rho)*expeta;
        c1 = log(lambda) + log(rho) + (rho-1)*log(Time) + eta
             - lambda*(Time**rho)*expeta;
        ll = (status=0)*c0 + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
        predict b1 out=bi;
        predict gamma(1+(1/rho))/(exp(eta)**(1/rho)) out=yhat;
```

```
        ods output ParameterEstimates = fixedsol;
        ods output hessian=hessian;
run;

proc sort data=asthma_for_local;
        by Patid;
run;

proc nlmixed data=asthma_for_local tech=quanew qpoints=50
                maxit=1000 hess start;
        bounds lambda > 0;
        parms /data=fixedsol;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = log(rho)-lambda*(Time**rho)*expeta;
        c1 = log(lambda) + log(rho) + (rho-1)*log(Time)
              + eta - lambda*(Time**rho)*expeta;
        ll = (status=0)*c0 + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
        by Patid;
        ods output ParameterEstimates = gradientid;
        ods output hessian=hessiaanid;
run;

data yhat;
        set yhat;
        keep Patid pred;
run;

proc iml;
        reset print;

        use asthma_for_local;
        labelx = {placebo ptime treatment ttime};
        labelz = {int};
        labely = {Time};
        read all var labelx into fixed;
        read all var labelz into random;
        read all var labely into resp;

        p=ncol(fixed);
        q=ncol(random);
        a=1;

        use yhat;
```

```
read all into yhat;
yhat= yhat[,2];

create yhat_program_1 from yhat;
append from yhat;

use fixedsol;
read all into fixedsol;
fixedpar= fixedsol[,1];

create fixedsol_program_1 from fixedpar;
append from fixedpar;

use hessian;
read all into L;
n_L = nrow(L);/*4*/

L= L[,2:1+nrow(L)];
L_inv=inv(L);

Lb= L[1:nrow(L)-q-a,1:nrow(L)-q-a];
Lb_inv=inv(Lb);

Ll= L[p-1:nrow(L)-a,p-1:nrow(L)-a];
Ll_inv=inv(Ll);

Ld= L[p:nrow(L)-a+1,p:nrow(L)-a+1];
Ld_inv=inv(Ld);

create L_1 from L;
append from L;

create Lb_1 from Lb;
append from Lb;

create Ll_1 from Ll;
append from Ll;

create Ld_1 from Ld;
append from Ld;

use gradientid;
read all into Delta;
delta_i= delta[,2];

create Delta_1 from delta_i;
append from delta_i;
```

```
use nfrec;
read all into nfrec;
id = nfrec[,1];
n_id=nrow(nfrec);

create nfrec_program_1 from id;
append from id;

begin = 1;
begin_b = 1;
begin_l = p-1;
begin_d = p+q-1;

do s=1 to n_id ;
        end = begin+p+q-2;
        end_b = begin_b+p-3;
        end_l = begin_l;
        end_d = begin_d;

        Ci = 2#delta_i[begin:end, ]`*L_inv
             *delta_i[begin:end, ];
        Ci_b = 2#delta_i[begin_b:end_b, ]`*Lb_inv
               *delta_i[begin_b:end_b, ];
        Ci_l = 2#delta_i[begin_l:end_l, ]`*Ll_inv
               *delta_i[begin_l:end_l, ];
        Ci_d = 2#delta_i[begin_d:end_d, ]`*Ld_inv
               *delta_i[begin_d:end_d, ];

        begin=end+1;
        begin_b=end_b+q+2;
        begin_l=end_l+p;
        begin_d=end_l+p;

        C_i = C_i//Ci;
        Cib = Cib//Ci_b;
        Cil = Cil//Ci_l;
        Cid = Cid//Ci_d;

index=index//s;
end;

begin = 1;

do s=1 to n_id ;
        ni = nfrec[s,2];
```

```
                end=begin+ni-1;

                fixedi = fixed[begin:end,];
                randomi = random[begin:end,];
                respi = resp[begin:end];
                yhati = yhat[begin:end];
                residi = respi-yhati;

                begin = end +1;

                rri = sqrt(trace(residi*residi`));
                xxi = sqrt(trace(fixedi*fixedi`));


                probnorm_rri = probnorm_rri//rri;
                probnorm_xxi = probnorm_xxi//xxi;


        end;

        out=index||C_i||Cib||Cil||Cid||probnorm_rri||probnorm_xxi;
        varnames = {'index' 'C_i' 'Ci_b' 'Ci_l' 'Ci_d'
                    '||rri||' '||xxi||'};

        create outdata_1 from out [colname= varnames];
        append from out;

        close fixedsol;
        close hessian;
        close nfrec;
        close asthma_for_local;
        close gradientid;

quit;

/*************************************************************
Local Influence for Exponential-Gamma-Normal Model
*************************************************************/

data asthma_for_local;
        set asthma;
        count + 1;
        by Patid;
        if first.Patid then count = 1;
run;

data asthma_for_local;
```

```
        set asthma_for_local;
        int=1;
        placebo = (Drug=0);
        treatment = (Drug=1);
        ptime = placebo*count;
        ttime = treatment*count;
run;

proc freq data=asthma_for_local noprint;
        tables Patid /out=nfrec;
run;

proc sort data=asthma_for_local;
        by Patid;
run;

/* Combined model with censoring */
proc nlmixed data=asthma_for_local tech=quanew qpoints=50
              maxit=1000 hess start;
        bounds lambda > 0, alpha > 0;
        parms Beta_0=-3 Beta_1=-0.11 lambda = 1 alpha=3.3
              sigma=1;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))
             **alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
        predict b1 out=bi;
        predict (gamma(alpha-(1/rho))*gamma(1+(1/rho)))
                /(gamma(alpha)*((exp(eta)/alpha)**(1/rho)))
                out=yhat;
        ods output ParameterEstimates = fixedsol;
        ods output hessian=hessian;
run;

proc sort data=asthma_for_local;
        by Patid;
run;

proc nlmixed data=asthma_for_local tech=quanew qpoints=50
                  maxit=1000 hess start;
```

```
        bounds lambda > 0, alpha > 0;
        parms /data=fixedsol;
        rho=1;
        eta = Beta_0 + Beta_1*(Drug=1) + b1;
        expeta = exp(eta);
        c0 = 1/((1 + lambda*expeta*(Time**rho)*(1/alpha))
             **alpha);
        c1 = log(lambda) + log(rho) + (alpha+1)*log(alpha)
              + (rho-1)*log(Time) + eta
              - (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
        ll = (status=0)*log(c0) + (status=1)*c1;
        model Time ~ general(ll);
        random b1 ~ normal(0, sigma**2) subject=Patid;
        by Patid;
        ods output ParameterEstimates = gradientid;
        ods output hessian=hessiaanid;
run;

data yhat;
        set yhat;
        keep Patid pred;
run;

proc iml;
        reset print;

        use asthma_for_local;
        labelx = {placebo ptime treatment ttime};
        labelz = {int};
        labely = {Time};
        read all var labelx into fixed;
        read all var labelz into random;
        read all var labely into resp;

        p=ncol(fixed);
        q=ncol(random);
        a=1;

        use yhat;
        read all into yhat;
        yhat= yhat[,2];

        create yhat_program_1 from yhat;
        append from yhat;

        use fixedsol;
        read all into fixedsol;
```

```
fixedpar= fixedsol[,1];

create fixedsol_program_1 from fixedpar;
append from fixedpar;

use hessian;
read all into L;
n_L = nrow(L);

L= L[,2:1+nrow(L)];
L_inv=inv(L);

Lb= L[1:nrow(L)-q-a-1,1:nrow(L)-q-a-1];
Lb_inv=inv(Lb);

Ll= L[p-1:nrow(L)-a-1,p-1:nrow(L)-a-1];
Ll_inv=inv(Ll);

La= L[p:nrow(L)-a,p:nrow(L)-a];
La_inv=inv(La);

Ld= L[p+q:nrow(L),p+q:nrow(L)];
Ld_inv=inv(Ld);

create L_1 from L;
append from L;

create Lb_1 from Lb;
append from Lb;

create Ll_1 from Ll;
append from Ll;

create La_1 from La;
append from La;

create Ld_1 from Ld;
append from Ld;

use gradientid;
read all into Delta;
delta_i= delta[,2];

create Delta_1 from delta_i;
append from delta_i;

use nfrec;
```

```
read all into nfrec;
id = nfrec[,1];
n_id=nrow(nfrec);

create nfrec_program_1 from id;
append from id;

begin = 1;
begin_b = 1;
begin_l=p-1;
begin_a = p;
begin_d = p+q;

do s=1 to n_id ;
        end=begin+p+q+a-2;
        end_b=begin_b+1;
        end_l=begin_l;
        end_a=begin_a;
        end_d=begin_d;

        Ci = 2#delta_i[begin:end, ]`*L_inv
             *delta_i[begin:end, ];
        Ci_b = 2#delta_i[begin_b:end_b, ]`*Lb_inv
               *delta_i[begin_b:end_b, ];
        Ci_l = 2#delta_i[begin_l:end_l, ]`*Ll_inv
               *delta_i[begin_l:end_l, ];
        Ci_a = 2#delta_i[begin_a:end_a, ]`*La_inv
               *delta_i[begin_a:end_a, ];
        Ci_d = 2#delta_i[begin_d:end_d, ]`*Ld_inv
               *delta_i[begin_d:end_d, ];

        begin=end+1;
        begin_b=end_b+q+a+2;
        begin_l=end_l+a+p;
        begin_a=end_a+p+1;
        begin_d=end_d+p+q;

        C_i = C_i//Ci;
        Cib = Cib//Ci_b;
        Cil = Cil//Ci_l;
        Cia = Cia//Ci_a;
        Cid = Cid//Ci_d;

index=index//s;
end;

begin = 1;
```

```
        do s=1 to n_id ;
                ni = nfrec[s,2];

                end=begin+ni-1;

                fixedi = fixed[begin:end,];
                randomi = random[begin:end,];
                respi = resp[begin:end];
                yhati = yhat[begin:end];
                residi = respi-yhati;

                begin = end +1;

                rri = sqrt(trace(residi*residi`));
                xxi = sqrt(trace(fixedi*fixedi`));

                probnorm_rri = probnorm_rri//rri;
                probnorm_xxi = probnorm_xxi//xxi;

        end;

        out=index||C_i||Cib||Cil||Cia||Cid||probnorm_rri
            ||probnorm_xxi;
        varnames = {'index' 'C_i' 'Ci_b' 'Ci_l' 'Ci_a' 'Ci_d'
                    '||rri||' '||xxi||'};

        create outdata_1 from out [colname= varnames];
        append from out;

        close fixedsol;
        close hessian;
        close nfrec;
        close asthma_for_local;
        close gradientid;

quit;
```

**LEUVEN STATISTICS RESEARCH CENTRE**
Celestijnenlaan 200B, bus 5307
3001 HEVERLEE, BELGIË
tel. + 32 16 32 22 42
fax + 32 16 32 28 31
www.kuleuven.be